Multiomics Analyses of HNF4α Protein Domain Function during Human Pluripotent Stem Cell Differentiation

Wang, Yu; Tatham, Michael H.; Schmidt-Heck, Wolfgang; Swann, Carolyn; Singh-Dolt, Karamjit; Meseguer-Ripolles, Jose

Published in:
iScience

DOI:
10.1016/j.isci.2019.05.028

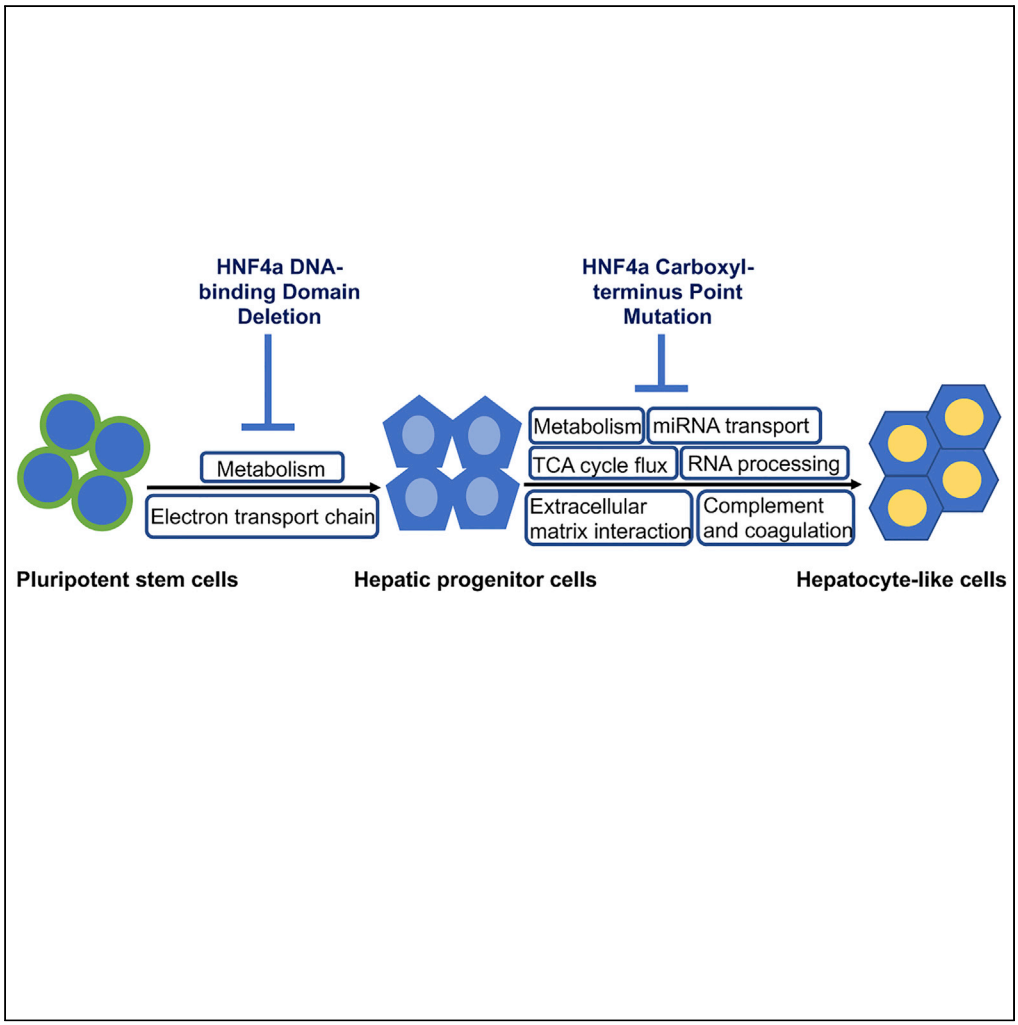
Publication date:
2019

iScience

CellPress

Article

Multiomics Analyses of HNF4α Protein Domain Function during Human Pluripotent Stem Cell Differentiation



Yu Wang, Michael H. Tatham, Wolfgang Schmidt-Heck, ..., Patricio Godoy, Ronald T. Hay, David C. Hay

davehay@talktalk.net

HIGHLIGHTS

Deletion of HNF4α's DNA-binding domain blocks hepatic progenitor formation

HNF4α's C-terminal SUMO consensus motif plays roles in hepatocyte specification

Multiomics analyses reveal various processes regulated by HNF4α during differentiation

Wang et al., iScience 16, 206–217
June 28, 2019
https://doi.org/10.1016/j.isci.2019.05.028

iScience

CellPress

Article

Multiomics Analyses of HNF4α Protein Domain Function during Human Pluripotent Stem Cell Differentiation

Yu Wang,[1] Michael H. Tatham,[2] Wolfgang Schmidt-Heck,[3] Carolyn Swann,[4] Karamjit Singh-Dolt,[1] Jose Meseguer-Ripolles,[1] Baltasar Lucendo-Villarin,[1] Tilo Kunath,[1] Timothy R. Rudd,[4] Andrew J.H. Smith,[1] Jan G. Hengstler,[5] Patricio Godoy,[5] Ronald T. Hay,[2] and David C. Hay[1,6,*]

SUMMARY

During mammalian development, liver differentiation is driven by signals that converge on multiple transcription factor networks. The hepatocyte nuclear factor signaling network is known to be essential for hepatocyte specification and maintenance. In this study, we have generated deletion and point mutants of hepatocyte nuclear factor-4alpha (HNF4α) to precisely evaluate the function of protein domains during hepatocyte specification from human pluripotent stem cells. We demonstrate that nuclear HNF4α is essential for hepatic progenitor specification, and the introduction of point mutations in HNF4α's Small Ubiquitin-like Modifier (SUMO) consensus motif leads to disrupted hepatocyte differentiation. Taking a multiomics approach, we identified key deficiencies in cell biology, which included dysfunctional metabolism, substrate adhesion, tricarboxylic acid cycle flux, microRNA transport, and mRNA processing. In summary, the combination of genome editing and multiomics analyses has provided *valuable* insight into the diverse functions of HNF4α during pluripotent stem cell entry into the hepatic lineage and during hepatocellular differentiation.

INTRODUCTION

Directed differentiation of human pluripotent stem cells offers robust systems to study gene function during human development. Defined and efficient human hepatic differentiation systems have been developed and automated, allowing detailed mechanistic studies to be performed (Hay et al., 2007, 2008; Lucendo-Villarin et al., 2017; Meseguer-Ripolles et al., 2018; Rashidi et al., 2018; Si-Tayeb et al., 2010; Sullivan et al., 2010; Szkolnicka et al., 2014). These models have been sophisticated further using modern genome editing systems. Recently, pluripotent stem cells with an albumin reporter system have facilitated the discovery of key players involved in hepatocyte maturation (Li et al., 2018). In our study, the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) endonuclease nickase was employed for hepatocyte nuclear factor-4alpha (HNF4α) genome editing in human pluripotent stem cells.

HNF4α is a highly conserved transcription factor of the nuclear receptor superfamily (Sladek et al., 1990). It is a modular protein containing five functional regions. Regions A and B contain the N-terminal transactivation domain (AD-1), region C is a highly conserved zinc finger DNA-binding domain (DBD), region D is a flexible hinge, region E is a multifunctional ligand-binding domain containing the second transactivation domain (AD-2), and the repression region is located in region F (Lau et al., 2018). Both transactivation domains are important to HNF4α-driven gene transcription (Dhe-Paganon et al., 2002). To understand which regions of the HNF4α protein are required for directing cell specification and maturation, we used CRISPR-Cas9-based genome editing technology to modify the DBD in the amino terminus, and AD-2 in the carboxy terminus.

Previous genome-wide location analysis revealed that HNF4α was bound to ~12% of genes in human hepatocytes, which was greater than other hepatocyte nuclear factors (Odom et al., 2004). Notably, 42% of the actively transcribed genes occupied by RNA polymerase II were bound by HNF4α, demonstrating its central role in hepatocyte biology (Odom et al., 2004). More recent researches showed that HNF4α was essential to the formation of hepatic progenitors during human pluripotent stem cell differentiation (DeLaForest et al., 2011, 2018). In these studies, HNF4α was required for the recruitment of RNA polymerase II to genes that were specifically expressed at the hepatic progenitor stage (DeLaForest et al., 2018).

[1]Medical Research Council Centre for Regenerative Medicine, University of Edinburgh, 5 Little France Drive, Edinburgh, Scotland EH16 4UU, UK

[2]Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dundee DD1 5EH, UK

[3]Leibniz Institute for Natural Product Research and Infection Biology eV-Hans-Knoll Institute, Jena, Germany

[4]National Institute for Biological Standards and Control (MHRA), Blanche Lane, South Mimms, Hertfordshire EN6 3QG, UK

[5]IfADo-Leibniz Research Centre for Working Environment and Human Factors at the Technical University Dortmund, Dortmund, Germany

[6]Lead Contact

*Correspondence: davehay@talktalk.net

https://doi.org/10.1016/j.isci.2019.05.028

Given its potent activity, HNF4α is regulated at numerous levels, which include nuclear receptor interaction, microRNA (miRNA) regulation, and post-translational modification (Lau et al., 2018; Ramamoorthy et al., 2012; Yokoyama et al., 2011). A number of post-translational modifications have been shown to regulate HNF4α protein stability, function, and subcellular localization (Jiang et al., 1997; Ktistaki et al., 1995; Viollet et al., 1997; Yokoyama et al., 2011; Zhou et al., 2012). In addition, recent studies highlighted that HNF4α protein stability is regulated post-translationally through interaction with heat shock protein 90 β interaction (Jing et al., 2017). In these studies, we were particularly interested in HNF4α's post-translational modification by the Small Ubiquitin-like Modifier (SUMO) at a consensus motif in the AD-2 domain of the carboxy terminus (Zhou et al., 2012).

This study examined HNF4α protein domain function during hepatoblast and hepatocyte specification from pluripotent stem cells. Isogenic pluripotent stem cell lines, which possessed a truncation or point mutations in HNF4α, were created and compared with a wild-type (WT) line. Pluripotent stem cells that possessed the DNA-binding domain deletion mutant of HNF4α displayed defects in endoderm and hepatic progenitor specification, whereas HNF4α point mutants failed to form functional hepatocytes, with defects in cell metabolism, adhesion, tricarboxylic acid (TCA) cycle flux, miRNA transport, and mRNA processing detected.

RESULTS

Generation and Characterization of *HNF4α*-Edited Pluripotent Stem Cell Lines

To reduce potential off-target effects, we used paired CRISPR-Cas9 nickases to edit *HNF4α* (Ran et al., 2013). Two pairs of Cas9 nickases were utilized to delete the consensus exon 2 in *HNF4α* (Figure 1A, top panel). The PCR product from amplification of the targeted region in a homozygous deletion mutant cell line was smaller than the WT control (Figure 1A, middle panel). There was a 540- and 541-bp deletion in each allele, respectively, which was confirmed by sequencing (Figure 1A, bottom panel). In parallel, two point mutations (K365R and D367A) were introduced into the SUMOylation consensus motif in the C terminus of HNF4α using Cas9 nickase and a piggyBac-based targeting vector (Zhou et al., 2012; Yusa, 2013). PCR genotyping followed by sequencing identified the insertion of the selection cassette in the targeted clones. The piggyBac repeats were inserted between 5′- and 3′-homology arms (Figure 1B, middle panel). Post removal of the selection cassette, sequencing results confirmed the seamless editing of this locus. In the point mutated clones, two point mutations were introduced into the *HNF4α* gene (AAG to AGG, K365R; GAC to GCC, D367A). Four synonymous mutations were also introduced to allow the integration of piggyBac (TTAA site) and to disrupt the protospacer-adjacent motif (PAM) sequence for Cas9 nickases (Figure 1B, bottom panel).

Following genome editing, one WT, one homozygous deletion mutant (DBD Mut), and one point mutated clone (SUMO Mut) were expanded, differentiated, and characterized in detail. Similar to the WT clone, the DBD Mut and SUMO Mut clones possessed typical pluripotent stem cell morphology, more than 90% cells expressed NANOG and OCT4, as well as cell surface markers SSEA4 and TRA-1-60 (Figures 1C and 1D). See Table S7 for further details.

Hepatocyte Differentiation Was Perturbed in *HNF4α*-Edited Pluripotent Stem Cells

To study the effect of editing *HNF4α* in liver cells, we differentiated WT, DBD Mut, and SUMO Mut clones towards hepatic lineage using a stage-wise differentiation protocol (Meseguer-Ripolles et al., 2018; Wang et al., 2017). WT human embryonic stem cells transited through definitive endoderm stage (differentiation day 3) (Figures S1A and S1B) and then gave rise to polygonal hepatic progenitors (differentiation day 9) (Figure 2A). In contrast, the DBD Mut cells displayed cytoplasmic HNF4α, prolonged endoderm differentiation, and failed to commit to hepatoblast lineage, demonstrated by the lack of alpha fetoprotein (AFP) and CCAAT enhancer-binding protein alpha (CEBPA) expression (Figures 2A and S1). In contrast, the SUMO Mut cells were comparable to the WT hepatoblast in terms of morphology, demonstrating nuclear HNF4α, albeit with reduced levels of HNF1α, AFP, and CEBPA expression (Figures 2A, 2C, and S1C).

Western bloting demonstrated that the DBD Mut hepatic progenitor cell expressed a truncated form of HNF4α, which was about 10 kDa smaller than the WT or SUMO Mut HNF4α (Figure 2B). Sequencing of HNF4α cDNA in the DBD Mut cells confirmed that exons 2 and 3 were skipped in these cells, resulting in the truncated form of HNF4α (Figure S2). A decrease in HNF4α expression was observed in all cell lines between days 9 and 16, which was consistent with previous studies in WT cells (Zhou et al., 2012) (Figure 2B).

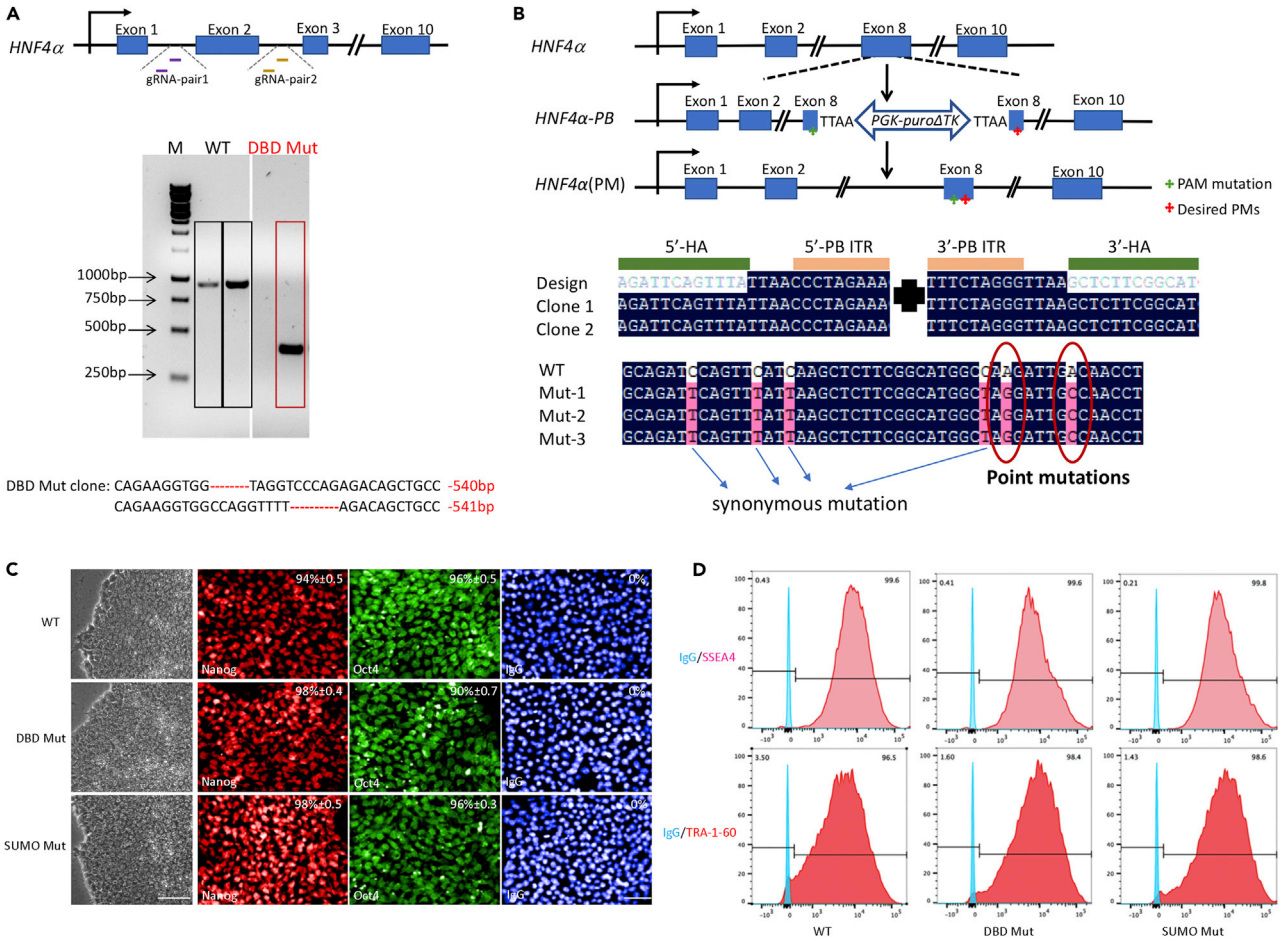


Figure 1. The Generation and Characterization of *HNF4α* Genome-Edited Cell Lines

(A) Two guide RNA pairs targeting introns 1 and 2 were used to delete exon 2 in *HNF4α* (top panel). A homozygous deletion clone was identified by amplifying the targeted region (middle panel). Sequencing confirmed the deletion mutant (DBD Mut) clone had a 540/541-bp deletion in each allele (bottom panel). See Tables S5 and S6 for further details.

(B) A piggyBac-based targeting vector was used in combination with Cas9 nickases to introduce desired point mutations into *HNF4α* (top panel). The targeted clones incorporated the selection cassette (middle panel). This selection cassette is contained within the piggyBac transposon and consists of a positive-negative selection marker (puro-tk) expressed from a constitutively active promoter (PGK). Post excision of the transposon, the locus was modified seamlessly (bottom panel). PAM, protospacer-adjacent motif; HA, homology arm; PB, piggyBac, 5′-PB ITR and 3′-PB ITR are 5′ and 3′ piggyBac inverted terminal repeats flanked by the TTAA direct repeats. See Tables S5 and S6 for further details.

(C) Representative images of cellular morphology, immunofluorescences of NANOG and OCT4. One wild-type (WT) clone, one DBD Mut clone, and one point-mutated (SUMO Mut) clone was selected for characterization. IgG was used as a negative control. The percentage was calculated using four random fields of view. Scale bar, 100 μm for phase contrast and 50 μm for immunostaining images.

(D) Flow cytometry of SSEA4- and TRA-1-60-expressing cells in the WT, DBD Mut, and SUMO Mut clones. IgG was used as a negative control. N = 3 independent experiments.

In addition, real-time PCR confirmed higher levels of HNF4α mRNA in the DBD Mut progenitor cells than in the WT or SUMO Mut cells (Figure 2C). Despite this, HNF4α in the DBD Mut cells failed to transactivate HNF1α and AFP gene expression and yielded lower levels of transthyretin (*TTR*) (Figures 2A and 2C).

In addition, neither the DBD Mut nor the SUMO Mut cells could produce hepatocyte-like cells. Their morphologies were different from each other and the WT control cells (Figure 2D). Functionally, the DBD Mut cells had no detectable albumin or AFP secretion, or basal cytochrome P450 (CYP) 3A activity (Figure 2E), demonstrating failed hepatoblast and hepatocyte differentiation. The SUMO Mut cells secreted AFP at lower levels than the WT cells; however, they did not exhibit albumin secretion or basal CYP3A activity when hepatocyte specification was induced, demonstrating failed hepatocyte specification (Figure 2E).

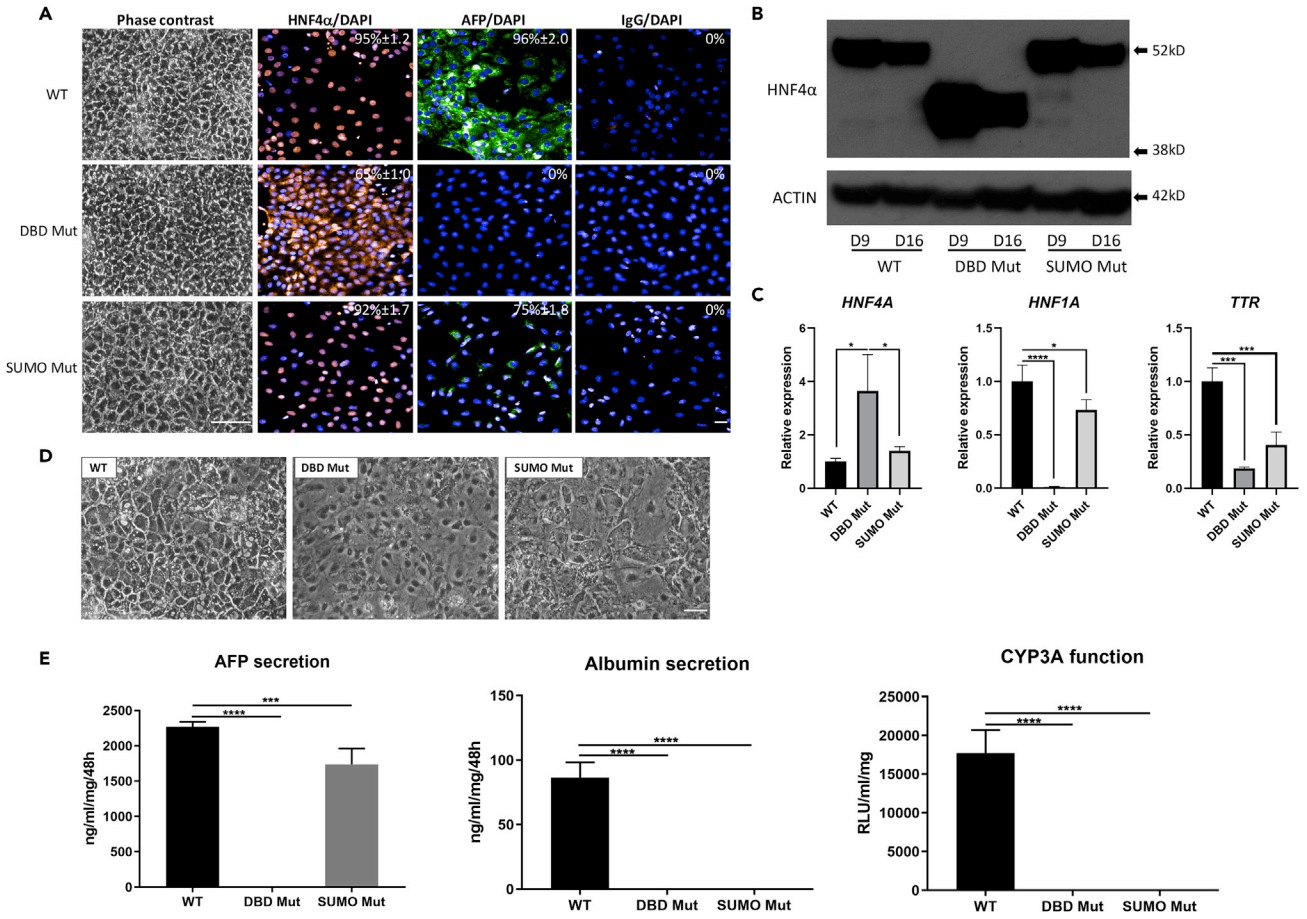


Figure 2. Hepatocyte Differentiation of *HNF4α* Genome-Edited Pluripotent Stem Cells

(A) Morphology and immunostaining of HNF4α and AFP in the differentiated cells at hepatic progenitor stage. Scale bar, 100 μm for phase contrast and 50 μm for immunostaining images.

(B) Western bloting for HNF4α in differentiated cells at days 9 (hepatic progenitor stage) and 16 (hepatocyte like cell). ACTIN was used as a loading control. See Table S7 for further details.

(C) Real-time PCR quantification of HNF4α, HNF1α, and transthyretin (TTR) mRNA levels in hepatic progenitor stage (day 9) cells. Data were normalized to the housekeeping gene ACTB and expressed relative to the WT cells. The results shown represent three biological replicates, and error bars represent SD. *p < 0.05, ***p < 0.001, ****p < 0.0001; one-way ANOVA with Tukey post hoc test. See Table S8 for further details.

(D) Morphology of differentiated cells at hepatocyte-like cell stage (differentiation day 18). Scale bar, 100 μm.

(E) ELISA quantification of AFP and albumin secretion in differentiated cells at hepatocyte-like cell stage and quantification of cytochrome P450 3A (CYP3A) activity. Data represented three biological replicates, and error bars represented SD. ***p < 0.001, ****p < 0.0001; one-way ANOVA with Tukey post hoc test.

Disruption of Cellular Bioenergetics in *HNF4α*-Edited Cells during Cell Specification

To gain an understanding of the DBD Mut and SUMO Mut cells at the hepatic progenitor stage (day 9), we performed non-targeted profiling of metabolites using ^{1}H nuclear magnetic resonance metabolomics (Patti et al., 2012). The metabolic rate for each metabolite was quantified by normalizing the signal intensity against the cell number at the detected time point. The unit for the metabolic rate was then recorded as rate of integral area per cell. Negative and positive rate values indicate consumption and production respectively. The DBD Mut and SUMO Mut cells had a significantly lower consumption rate for both glucose and pyruvate than the WT control (Figure 3A). This was confirmed by reduced production rate of lactate, formate, and acetate (Figure 3A) and suggests compromised glycolysis and pyruvate oxidation or cycling in both cell lines. In addition, the DBD Mut and SUMO Mut cells possessed a lower consumption rate of the essential amino acids threonine and tryptophan compared with the WT cells (Figure 3B). The SUMO Mut cells also demonstrated slower consumption of valine, methionine, and phenylalanine than both WT and DBD Mut cells (Figure 3B). Notably, the DBD Mut cells had the highest consumption rate of isoleucine among the three cell types (Figure S3). The WT cells had significantly increased production

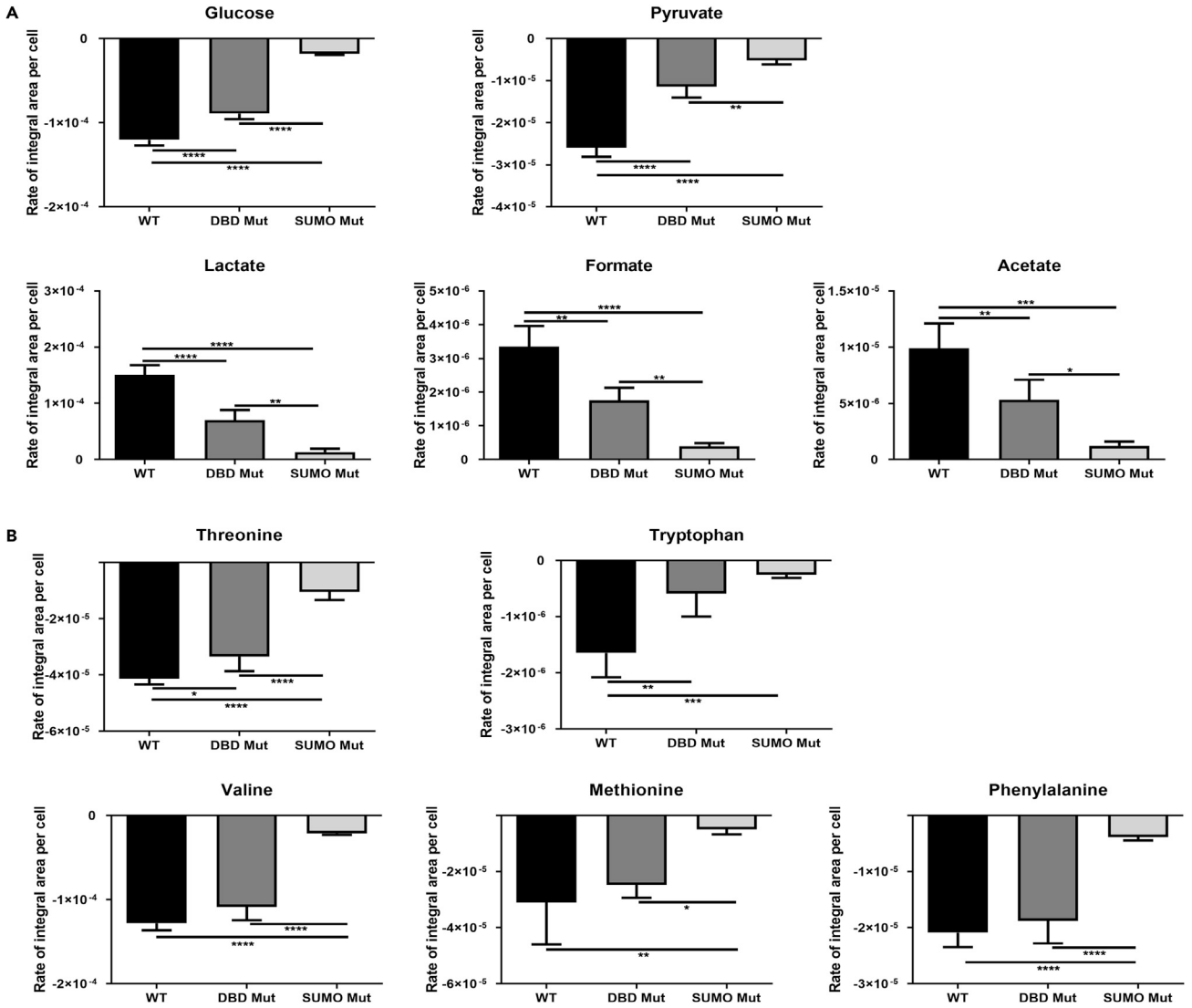


Figure 3. Metabolomic Studies in Hepatic Progenitor Stage Cells from WT and *HNF4α* Genome-Edited Pluripotent Stem Cells

(A) Metabolic rates of glucose, pyruvate, lactate, formate, and acetate in the WT, DBD Mut, and SUMO Mut cells.

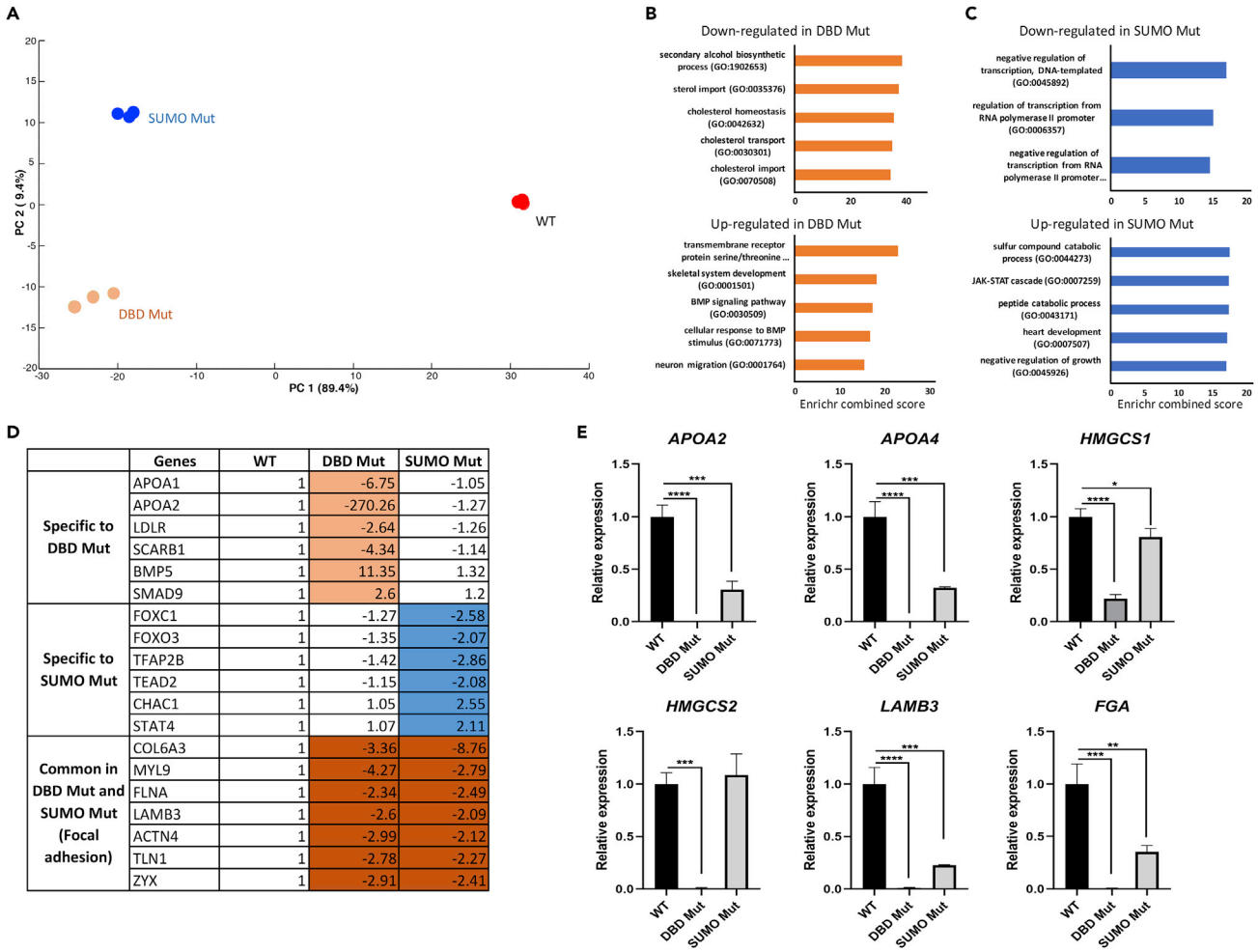
(B) Metabolic rates of essential amino acids threonine, tryptophan, valine, methionine, and phenylalanine in the WT, DBD Mut, and SUMO Mut cells.

Data represent three biological replicates, and error bars represent SD. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001; one-way ANOVA with Tukey post hoc test.

of alanine and consumption of tyrosine and t-methylhistidine than the other cell lines (Figure S3). These data demonstrated that amino acid metabolism and bioenergetics were disrupted in HNF4α DBD Mut and SUMO Mut cells.

Transcriptomic Analysis of WT and *HNF4α*-Edited Hepatic Progenitor Stage Cells

The cellular phenotypes we observed highlighted that proper HNF4α function was vital to cell metabolism and hepatocyte specification. To further understand the effect of *HNF4α* editing on hepatic progenitor cell biology, transcriptomic profiles were generated for WT, DBD Mut, and SUMO Mut cells at differentiation day 9. Principal-component analysis (PCA) of 1,000 genes with the highest variance demonstrated a clear difference among DBD Mut, SUMO Mut, and WT (Figures 4A and S4). A set of 1,310 differentially expressed genes (DEGs) was defined for further evaluation, which were either differential between the DBD Mut and WT cells or between the SUMO Mut and WT cells (p value <0.05; 2-fold threshold, false discovery rate adjusted; Table S1). Gene ontology (GO) enrichment analysis using the Enrichr web server (Chen et al.,



D

	Genes	WT	DBD Mut	SUMO Mut
Specific to DBD Mut	APOA1	1	-6.75	-1.05
	APOA2	1	-270.26	-1.27
	LDLR	1	-2.64	-1.26
	SCARB1	1	-4.34	-1.14
	BMP5	1	11.35	1.32
	SMAD9	1	2.6	1.2
Specific to SUMO Mut	FOXC1	1	-1.27	-2.58
	FOXO3	1	-1.35	-2.07
	TFAP2B	1	-1.42	-2.86
	TEAD2	1	-1.15	-2.08
	CHAC1	1	1.05	2.55
	STAT4	1	1.07	2.11
Common in DBD Mut and SUMO Mut (Focal adhesion)	COL6A3	1	-3.36	-8.76
	MYL9	1	-4.27	-2.79
	FLNA	1	-2.34	-2.49
	LAMB3	1	-2.6	-2.09
	ACTN4	1	-2.99	-2.12
	TLN1	1	-2.78	-2.27
	ZYX	1	-2.91	-2.41

Figure 4. Transcriptomic Analysis of Hepatic Progenitor Stage Cells from WT and *HNF4α* Genome-Edited Pluripotent Stem Cells

(A) Principal-component analysis of the three types of cells (n = 3 replicates) based on 1,000 genes with the highest variance.

(B) Top five enriched biological processes of the significantly down- or up-regulated genes specific in the DBD Mut cells. Pathway enrichment was analyzed using Enrichr web server.

(C) Top five enriched biological processes of the significantly down- or up-regulated genes unique to the SUMO Mut cells. Pathway enrichment was analyzed using Enrichr web server.

(D) Fold changes of representative genes in the DBD Mut and SUMO Mut cells compared with the WT control from the microarray analysis. The gene expression levels have been log2 transformed.

(E) Real-time PCR results of representative genes *APOA2*, *APOA4*, *HMGCS1*, *HMGCS2*, laminin subunit beta 3 (*LAMB3*), and fibrinogen alpha chain (*FGA*). Data were normalized to the housekeeping gene ACTB and expressed relative to the WT cells. The results shown represent three biological replicates, and error bars represent SD. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001; one-way ANOVA with Tukey post hoc test. See Table S8 for further details.

2013; Kuleshov et al., 2016) showed that the 452 down-regulated genes unique to DBD Mut cells were enriched in biological processes such as "sterol import (GO:0035376)" and "cholesterol homeostasis (GO:0042632)" (Figure 4B and Table S2). Representative genes included apolipoprotein A1 (*APOA1*), *APOA2*, and low-density lipoprotein receptor (*LDLR*) (Figure 4D and Table S2). The expression of a number of significantly down-regulated genes was confirmed by real-time PCR results, including *APOA2*, *APOA4*, and 3-hydroxy-3-methylglutaryl-CoA synthase 1 and 2 (*HMGCS1* and *HMGCS2*) (Figure 4E). We also detected 215 unique up-regulated genes in the DBD Mut cells. Those were most enriched in pathways like "transmembrane receptor protein serine/threonine kinase signaling pathway (GO:0007178)" and "skeletal system development" (Figure 4B and Table S2). The representative genes included bone morphogenetic protein 5 (*BMP5*) and *SMAD9* (Figure 4D and Table S1). Interestingly, all up-regulated genes in the DBD Mut cells compared with the WT control showed features for ectoderm differentiation (Table S2). When combined, these specific DEGs in the Del Mut cells were most enriched in cholesterol homeostasis,

extracellular matrix (ECM) organization, alpha-amino acid catabolic process, and cytoskeleton organization regulation (Table S2). Taken together, the HNF4α DBD domain is essential for the expression of genes important for cell metabolism, ECM, and cytoskeleton organization during hepatoblast specification.

We also analyzed the specific DEGs in the SUMO Mut cells to understand the effect of modifying the HNF4α SUMO motif. We detected 260 down-regulated genes specific to the SUMO Mut cells. Those were enriched for biological processes including "negative regulation of transcription (GO:0045892)" and "regulation of transcription from RNA polymerase II promoter (GO:0006357)" (Figure 4C and Table S2). Genes such as the forkhead box protein C1 and O3 (*FOXC1* and *FOXO3*) were significantly down-regulated (Figure 4D and Table S1). We also detected 141 up-regulated genes enriched for processes like "sulfur compound catabolic process (GO:0044273)" and "JAK-STAT cascade (GO:0007259)" (Figure 4C and Table S2). Representative genes included ChaC glutathione-specific gamma-glutamylcyclotransferase 1 (*CHAC1*) and signal transducer and activator of transcription 4 (*STAT4*) (Figure 4D and Table S1). Combining the significantly down- and up-regulated genes together, the most enriched biological process was negative regulation of gene transcription (Table S2). Collectively, the introduced point mutations in the HNF4α SUMO consensus motif most likely affected gene transcription necessary for successful hepatocyte specification.

We also studied transcriptomic differences between the different mutant forms of HNF4α; 584 genes were down-regulated in the DBD Mut cells when compared with the SUMO Mut. They were mainly enriched in cholesterol homeostasis and lipoprotein remodeling. There were also 242 genes up-regulated in the DBD Mut cells when compared with the SUMO Mut cells. These genes demonstrated a shift towards muscle cell differentiation (Tables S1 and S2).

Further analysis revealed that 169 DEGs were down-regulated in both DBD Mut and SUMO Mut cells when compared with the WT. By referring to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database, the focal adhesion pathway was the only one significantly (p value < 0.05) enriched (Table S2). Representative genes encoded proteins for the basement membrane (laminin subunit beta 3), actin binding (filamin A and actinin alpha 4), and ECM binding (collagen type VI alpha 3 chain and talin 1) (Figures 4D and 4E and Table S1). There were also 74 DEGs up-regulated in both the DBD Mut and SUMO Mut cells; however, no pathway was significantly enriched. Therefore focal adhesion alterations in both DBD Mut and SUMO Mut cells probably contributed to hepatoblast and hepatocyte formation.

Proteomic Analysis of WT and *HNF4α*-Edited Hepatic Progenitor Stage Cells

Following on from the metabolic and transcriptomic studies, we examined alterations in the proteome at differentiation day 9 from WT, DBD Mut, and SUMO Mut lines. Tryptic peptides were analyzed by LC-MS/MS, and the raw MS output data was processed using the MaxQuant platform and then analyzed in Perseus software (Tyanova et al., 2015, 2016). A total of 3,639 protein groups were evaluated across all cell types. PCA analysis showed that the DBD Mut cells were distinct from the WT and SUMO Mut cells (Figure 5A and S5A). Proteins with reduced expression unique to the DBD Mut cells (Table S3), compared with WT, were involved in metabolic processes and oxidative phosphorylation (Figure 5B and Table S4). Notably, 119 of the 486 less abundant proteins specifically in the DBD Mut cells were mitochondrial proteins (Table S3). A number of subunits of Complexes I, II, III, IV, and V from the respiratory electron transport chain were down-regulated in the DBD Mut cells (Figure 5B). These included the core subunits of Complex I NADH dehydrogenase [ubiquinone] flavoprotein 1 and 2 (NDUFV1 and NDUFV2) and NADH dehydrogenase [ubiquinone] iron-sulfur protein 3 (NDUFS3) (Figure 5B and Table S3). In addition, the most down-regulated proteins in the DBD Mut included mitochondrial HMGCS2 and glycine dehydrogenase (GLDC) (Table S3). HMGCS2 catalyzes the first step of ketogenesis, whereas GLDC is involved in glycine degradation. The decreased expression of HMGCS2 and GLDC was consistent at the mRNA level, together with other metabolic enzymes such as glutamate dehydrogenase 1 (GLUD1) and aldehyde dehydrogenase 5 family member A1 (ALDH5A1) (Figures 4E and S5B). Taken together, the HNF4α DBD domain is essential to mitochondrial function and cellular metabolism.

Proteins detected at reduced levels in the SUMO Mut cells were significantly enriched for ECM organization and lipoprotein metabolic pathways (Figure 5C and Table S4). A dense network of associations was found around heparan sulfate proteoglycan 2 (HSPG2) in STRING, which is a web resource of known and predicted protein-protein interactions (Szklarczyk et al., 2015). HSPG2 encodes the perlecan protein, a core component of basement membranes. HSPG2 formed strong associations with the thyroid hormone transporter TTR, apolipoproteins APOA2 and APOA4, and the endocytic receptor

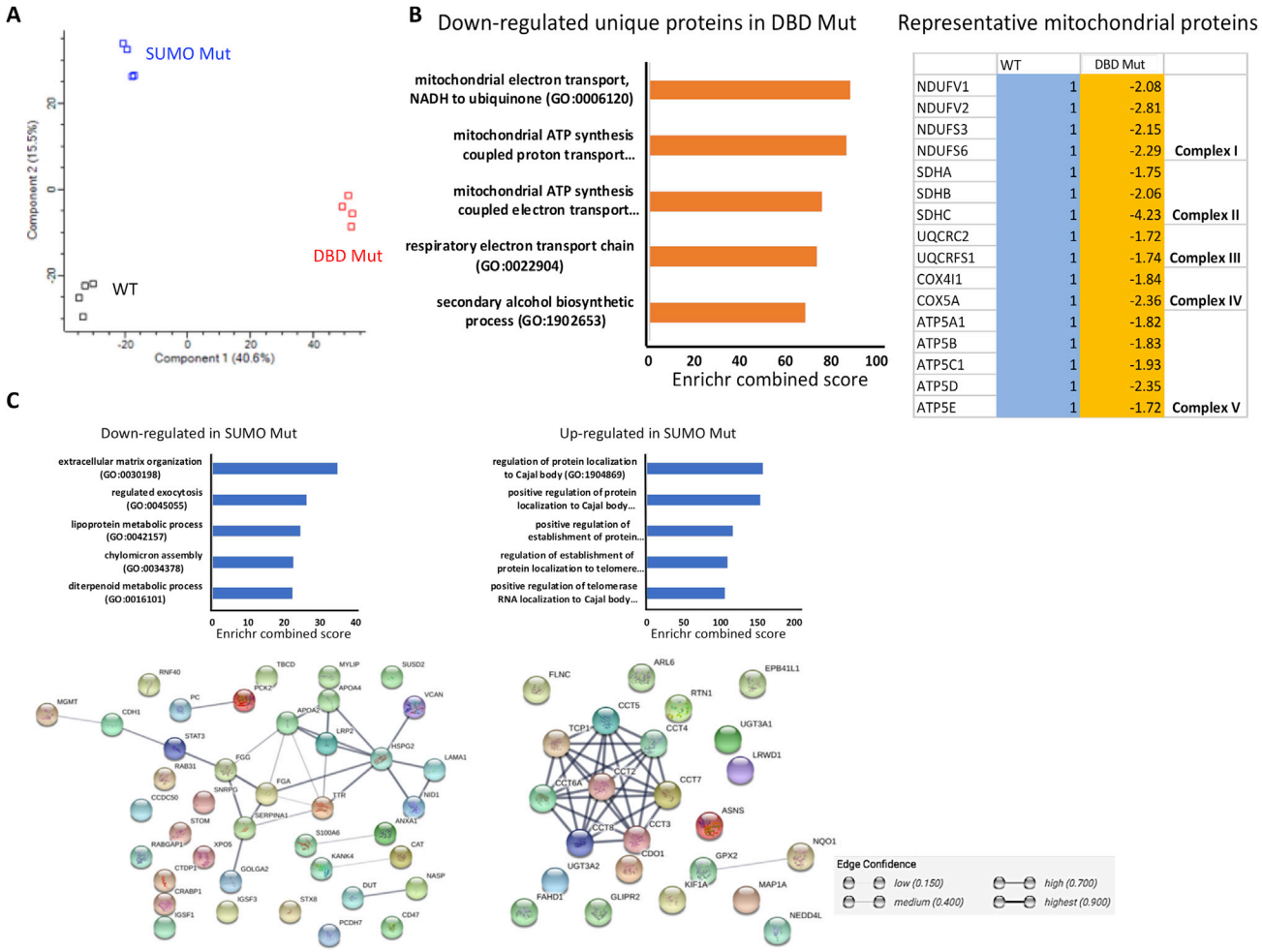


Figure 5. Proteomic Analysis of Hepatic Progenitor Stage Cells from WT and *HNF4α* Genome-Edited Pluripotent Stem Cells

(A) Principal component analysis of the WT, DBD Mut, and SUMO Mut cells at hepatic progenitor stage (n = 4 replicates).

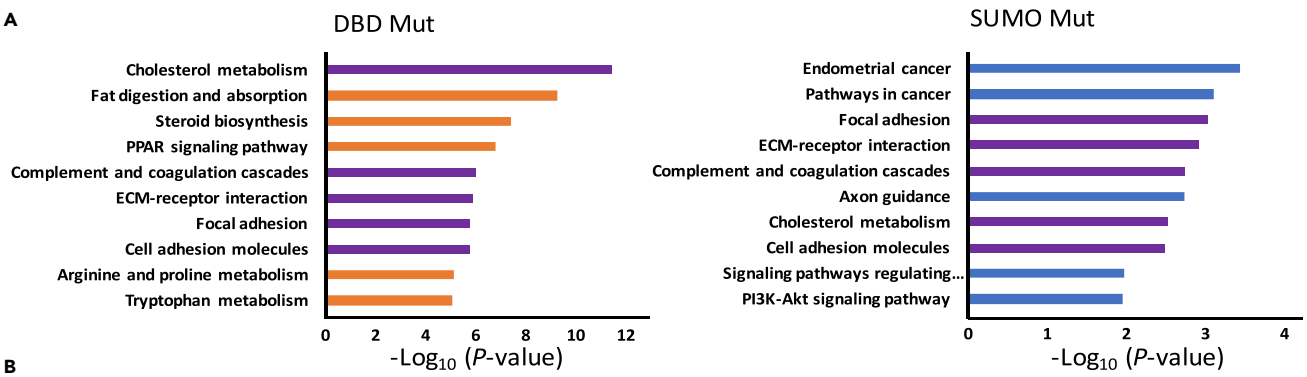
(B) Top five enriched biological processes of the significantly down-regulated proteins specifically in the DBD Mut cells and the representative down-regulated mitochondrial proteins in the DBD Mut cells. Pathway enrichment was analyzed using Enrichr server.

(C) Top five enriched biological processes and the protein-protein interaction networks of the significantly down- or up-regulated proteins in the SUMO Mut cells compared with the WT. Pathway enrichment was analyzed using Enrichr server. The protein-protein interaction networks were generated on STRING server. Network nodes represent proteins, and edges represent protein-protein associations. The bolder the edge, the higher the confidence.

low-density lipoprotein receptor-related protein 2 (LRP2), the basement membrane protein nidogen-1(NID1), ECM proteins laminin subunit alpha-1 (LAMA1) and versican (VCAN), and the alpha chain of the coagulation factor fibrinogen (FGA) (Figure 5C). Proteins that were more abundant in the SUMO Mut cells were mostly involved in the regulation of protein localization to Cajal body and telomeres (Figure 5C and Table S4). A dense network of associations was also formed by chaperonin-containing T-complex subunits (CCT1-8) (Figure 5C). Of note, SUMO Mut cells also contained differentially expressed proteins that were significantly altered for by at least 8-fold (Table S3). These proteins were involved in diverse cellular processes, including gluconeogenesis and TCA cycle flux (phosphoenolpyruvate carboxykinase 2 [PCK2]), as well as microRNA transport (Exportin 5). This suggested that the introduced point mutations in HNF4α not only affected ECM organization and cellular metabolism but also led to defects in miRNA transport.

There were 27 commonly down-regulated proteins in the DBD Mut and SUMO Mut cells when compared with the WT control. These proteins were most enriched in KEGG pathways such as complement and coagulation cascades, ECM-receptor interaction, pyruvate metabolism, and TCA cycle (Table S4). The

A



B

	Proteins	WT	DBD Mut	SUMO Mut
Respiratory electron transport chain	NDUFV1	1	-2.08	-1.78
	NDUFV2	1	-2.82	-1.96
	NDUFS3	1	-2.15	-1.52
	ATP5A1	1	-1.82	-1.24
	ATP5B	1	-1.83	-1.28
	ATP5C1	1	-1.93	-1.28
Amino acid metabolism	GLDC	1	-111.11	-1.54
	GLUD1	1	-8.1	-1.63
Ketogenesis	HMGCS2	1	-222.22	1.37
	HMGCS1	1	-13.57	1.02
TCA cycle flux	PCK2	1	-13.46	-10.49
	PC	1	-31.85	-3.64
Cholesterol metabolism	APOA2	1	-8.54	-12.03
	APOA4	1	-416.67	-3.91
Basement membrane	HSPG2	1	-27.47	-3.22
	NID1	1	-10.82	-23.7
Extracellular matrix	LAMA1	1	-4.15	-2.58
	VCAN	1	-3.87	-2.94
Coagulation factor	FGA	1	-161.29	-2.65
	FGG	1	-43.29	-3.64

Figure 6. Integrated Transcriptomic and Proteomic Analysis of Hepatic Progenitor Stage Cells Differentiated from WT and *HNF4α* Genome-Edited Pluripotent Stem Cells

(A) Top 10 enriched KEGG pathways in the DBD Mut cells or in the SUMO Mut cells from the overlayered analysis of both transcriptomic and proteomic datasets.

(B) Representative proteins and their fold changes in expression in the DBD Mut and SUMO Mut cells when compared with WT cells. The label-free quantification intensities have been log10 transformed.

representative proteins included pyruvate carboxylase (PC) and PCK2 (Figure S5C), which are involved in gluconeogenesis, pyruvate metabolism, and TCA cycle flux, supporting our metabolomic analysis.

Integrating Transcriptomics and Proteomics Datasets

To integrate information from both transcriptomics and proteomics experiments, we re-analyzed both datasets for the edited cells (DBD Mut/WT and SUMO Mut/WT) using a free web-based multiomics data visualization application named PaintOmics 3 (http://bioinfo.cipf.es/paintomics/) (Hernandez-de-Diego et al., 2018). In the DBD Mut cells, the most enriched KEGG pathway was cholesterol metabolism, followed by pathways in peroxisome proliferator-activated receptor signaling, ECM-receptor interaction, focal adhesion, and amino acid metabolism. Five of the top 10 enriched pathways in the DBD Mut cells were also in the top 10 significantly enriched pathways in the SUMO Mut cells. They were cholesterol metabolism, complement and coagulation cascades, ECM-receptor interaction, focal adhesion, and cell adhesion molecules (Figure 6A). Based on these analyses, we summed up the representative proteins that were significantly altered in the DBD Mut cells and in both DBD Mut and SUMO Mut cells (Figure 6B and Table S3). Taken together, the DBD Mut and SUMO Mut cells shared similar dysregulated genes and proteins that function in cellular metabolism and cell-ECM interaction. Meanwhile, the DBD Mut cells had disrupted respiratory electron transport chain and metabolism, highlighting the importance of the HNF4α DBD domain.

DISCUSSION

The hepatocyte nuclear factor network is vital for mammalian liver development and organ homeostasis (Odom et al., 2006; Watt et al., 2003). This network controls the expression of a large number of hepatic genes, which perform a broad range of functions (Battle et al., 2006; Bolotin et al., 2010; Odom et al., 2004). In the present study, we investigated HNF4α function during hepatic specification from human pluripotent stem cells. Without its DNA-binding domain, the truncated form of HNF4α failed to make hepatic progenitor cells, and displayed disruption in key processes such as the respiratory electron transport chain and cellular metabolism. Conversely, the introduction of point mutations, in the C-terminal SUMO consensus motif of HNF4α permitted hepatic progenitor commitment, but led to failed hepatocyte specification. Key genes involved in metabolism, TCA cycle flux, miRNA transport, cell-ECM interactions, mRNA processing, and coagulation cascades were implicated in this study.

Pluripotent stem cells rely heavily on aerobic glycolysis during self-renewal, but they switch quickly to oxidative phosphorylation (OXPHOS) during mesoderm and endoderm differentiation to meet the energy demand. Interestingly, during hepatic differentiation from human pluripotent stem cells, OXPHOS has been reported to decrease when cells reach the hepatic progenitor stage, followed by an increase during hepatocyte maturation (Hopkinson et al., 2017; Jing et al., 2018). In our study, we integrated omics analyses at the hepatic progenitor stage and found that ketogenesis, amino acid metabolism, TCA flux, and cholesterol metabolism pathways were largely down-regulated in mutated cells. We propose that these alterations were contributory factors that led to failed hepatocyte specification.

Previous studies in rodents have also shown that HNF4α-null embryonic livers display defects in glycogen synthesis (Parviz et al., 2003). In addition, glucose responsiveness was abolished in murine embryoid-body-derived visceral endoderm when HNF4α was deleted (Stoffel and Duncan, 1997). In this study, we demonstrate that HNF4α is critical to glucose utilization at the hepatic progenitor stage (Figure 3A). When HNF4α was truncated or point mutated, glucose and pyruvate utilization were compromised (Figures 3A and S3). Similarly, PC and PCK2, key enzymes in pyruvate cycling, were significantly down-regulated in genome-edited cells (Figure 6B). Notably, the SUMO Mut cells showed more pronounced metabolomic changes than the DBD Mut cells (Figure 3). The reason behind this is unknown, but it could be that the knockin mutations affected HNF4α's second transactivation domain (AD-2) leading to alterations in gene transcription.

Although the liver is regarded as one of the body's metabolic centers, less is known about the amino acid metabolism dynamics during early hepatogenesis. A recent study reported that L-valine is essential to murine liver bud growth and promotes the propagation of human hepatic progenitor organoids (Koike et al., 2017). This study did not investigate whether valine was required for human liver bud formation, but we note that valine consumption was significantly reduced in both DBD Mut and SUMO Mut cells at the hepatic progenitor stage. Although valine was not essential for hepatoblast specification in SUMO Mut cells, it is possible that valine was required for the specification of bipotent hepatic progenitors. However, further research is required to test this hypothesis.

Cell-ECM interaction also play a critical role in cellular differentiation. In the DBD Mut and SUMO Mut hepatic progenitor stage cells, we identified commonly down-regulated genes. They were enriched for focal adhesion and ECM organization pathways (Figures 4 and 6, Supplemental Information, Tables S2 and S4). Specific to the SUMO Mut cells, exportin-5 (XPO5) expression was reduced, potentially affecting precursor miRNA transport from the nucleus to the cytoplasm. We also observed down-regulation of genes that regulate gene transcription and mRNA processing, such as *FOXC1*, *FOXO3* (Figure 4D), and small nuclear ribonucleoprotein G (SNRPG) (Supplemental Information, Table S3). We believe that these data point to deficiencies in multiple cell biological processes that are instrumental in the hepatoblast to hepatocyte transition.

In summary, this study highlights the important role played by HNF4α during hepatic endoderm differentiation. System biology analyses revealed numerous potential regulatory functions for HNF4α. Those include the control of the respiratory electron transport chain, cell metabolism, pyruvate cycling, TCA cycle flux, miRNA transport, mRNA processing, and cell-ECM interaction.

Limitations of the Study

This study was based on an *in vitro*-directed differentiation system and was not tested *in vivo*.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.05.028.

ACKNOWLEDGMENTS

We thank Kosuke Yusa for sharing the pMCS-AAT_PB-PGKpuroTK, and pCMV-hyPBase plasmids. We thank Katharina Rochlitz for isolating RNA for microarray analysis, Asa L. Nordgren for contributing to the script used in metabolomics analysis, and Suresh Kushik for helpful discussions in genome editing. D.C.H. lab is supported by an award from the Chief Scientist Office (TC/16/37) and the Human Consortium (EU FP7 EC 602757). Y.W. was supported by a PhD scholarship funded by the Chinese Scholarship Council and the University of Edinburgh. This work is partially supported by Cancer Research UK Grant (C434/A13067) to M.H.T., a Wellcome Trust Senior Investigator award (098391/Z/12/Z) to R.T.H., and BMBF project StemCellNet (01EK1604A) to P.G. with some contributions of the projects Liver Simulator (BMBF, 031A355A), DILI (BMBF, 031L0074F), LiSyM (BMBF, 031Loo45), LivSysTransfer (BMBF, 0101-31Q0517), InnoSysTox (BMBF/EU, 031L0021A), WISP1 (DFG, Go1987/2-1), IL-15 (GO1987/3-1), DEEP (BMBF, 01KU1216), and EUToxRisk (EU, no. 681002).

This paper is based on independent research commissioned and funded by the NIHR Policy Research Program (NIBSC Regulatory Science Research Unit). The views expressed in the publication are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health, ''arms'' length bodies, or other government departments.

AUTHOR CONTRIBUTIONS

Y.W. and D.C.H. conceived the study. Y.W. performed the experiments, analyzed the data, and interpreted the omics analysis. M.H.T. and R.T.H. acquired and analyzed the proteomics data. W.S.-H, P.G., and J.G.H. acquired and analyzed the microarray data. C.S and T.R.R acquired and analyzed the metabolomics data. K.S.-D., T.K., and A.J.H.S. contributed to manuscript writing, the construction of the targeting vector and genome editing design. J.M.-R. and B.L.-V. helped with cell culture and data interpretation. Y.W. and D.C.H. wrote the paper, with input from all authors.

DECLARATION OF INTERESTS

D.C.H is a co-founder, shareholder, and director of Stemnovate Limited and Higher Steaks Limited.

Received: January 16, 2019
Revised: April 8, 2019
Accepted: May 21, 2019
Published: June 28, 2019

REFERENCES

Battle, M.A., Konopka, G., Parviz, F., Gaggl, A.L., Yang, C., Sladek, F.M., and Duncan, S.A. (2006). Hepatocyte nuclear factor 4 orchestrates expression of cell adhesion proteins during the epithelial transformation of the developing liver. Proc. Natl. Acad. Sci. U S A *103*, 8419–8424.

Bolotin, E., Liao, H., Ta, T.C., Yang, C., Hwang-Verslues, W., Evans, J.R., Jiang, T., and Sladek, F.M. (2010). Integrated approach for the identification of human hepatocyte nuclear factor 4α target genes using protein binding microarrays. Hepatology *51*, 642–653.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

DeLaForest, A., Nagaoka, M., Si-Tayeb, K., Noto, F.K., Konopka, G., Battle, M.A., and Duncan, S.A. (2011). HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. Development *138*, 4143–4153.

DeLaForest, A., Di Furio, F., Jing, R., Ludwig-Kubinski, A., Twaroski, K., Urick, A., Pulakanti, K., Rao, S., and Duncan, S. (2018). HNF4A regulates the formation of hepatic progenitor cells from human iPSC-derived endoderm by facilitating efficient recruitment of RNA Pol II. Genes (Basel) *10*, 21.

Dhe-Paganon, S., Duda, K., Iwamoto, M., Chi, Y.I., and Shoelson, S.E. (2002). Crystal structure of the HNF4α ligand binding domain in complex with endogenous fatty acid ligand. J. Biol. Chem. *277*, 37973–37976.

Hay, D.C., Zhao, D., Ross, A., Mandalam, R., Lebkowski, J., and Cui, W. (2007). Direct differentiation of human embryonic stem cells to hepatocyte-like cells exhibiting functional activities. Cloning Stem Cells *9*, 51–62.

Hay, D.C., Fletcher, J., Payne, C., Terrace, J.D., Gallagher, R.C.J., Snoeys, J., Black, J.R., Wojtacha, D., Samuel, K., Hannoun, Z., et al. (2008). Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA

and Wnt3a signaling. Proc. Natl. Acad. Sci. U S A *105*, 12301–12306.

Hernandez-de-Diego, R., Tarazona, S., Martinez-Mira, C., Balzano-Nogueira, L., Furio-Tari, P., Pappas, G.J., and Conesa, A. (2018). PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. BioRxiv, 1–7, https://www.biorxiv.org/content/early/2018/03/13/281295.full.pdf.

Hopkinson, B.M., Desler, C., Kalisz, M., Vestentoft, P.S., Juel Rasmussen, L., and Bisgaard, H.C. (2017). Bioenergetic changes during differentiation of human embryonic stem cells along the hepatic lineage. Oxid. Med. Cell. Longev. *2017*, 5080128.

Jiang, G., Nepomuceno, L., Yang, Q., and Sladek, F.M. (1997). Serine/threonine phosphorylation of orphan receptor hepatocyte nuclear factor 4. Arch. Biochem. Biophys. *340*, 1–9.

Jing, R., Duncan, C.B., and Duncan, S.A. (2017). A small-molecule screen reveals that HSP90β promotes the conversion of induced pluripotent stem cell-derived endoderm to a hepatic fate and regulates HNF4A turnover. Development *144*, 1764–1774.

Jing, R., Corbett, J.L., Cai, J., Beeson, G.C., Beeson, C.C., Chan, S.S., Dimmock, D.P., Lazcares, L., Geurts, A.M., Lemasters, J.J., et al. (2018). A screen using iPSC-derived hepatocytes reveals NAD + as a potential treatment for mtDNA depletion syndrome. Cell Rep. *25*, 1469–1484.e5.

Koike, H., Zhang, R.-R., Ueno, Y., Sekine, K., Zheng, Y.-W., Takebe, T., and Taniguchi, H. (2017). Nutritional modulation of mouse and human liver bud growth through a branched-chain amino acid metabolism. Development *144*, 1018–1024.

Ktistaki, E., Ktistakis, N.T., Papadogeorgaki, E., and Talianidis, I. (1995). Recruitment of hepatocyte nuclear factor 4 into specific intranuclear compartments depends on tyrosine phosphorylation that affects its DNA-binding and transactivation potential. Proc. Natl. Acad. Sci. U S A *92*, 9876–9880.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Lau, H.H., Ng, N.H.J., Loo, L.S.W., Jasmen, J.B., and Teo, A.K.K. (2018). The molecular functions of hepatocyte nuclear factors – in and beyond the liver. J. Hepatol. *68*, 1033–1048.

Li, S., Li, M., Liu, X., Yang, Y., Wei, Y., Chen, Y., Qiu, Y., Zhou, T., Feng, Z., Ma, D., et al. (2018). Genetic and chemical screenings identify HDAC3 as a key regulator in hepatic differentiation of human pluripotent stem cells. Stem Cell Reports *11*, 22–31.

Lucendo-Villarin, B., Filis, P., Swortwood, M.J., Huestis, M.A., Meseguer-Ripolles, J., Cameron, K., Iredale, J.P., O'Shaughnessy, P.J., Fowler, P.A., and Hay, D.C. (2017). Modelling foetal exposure to maternal smoking using hepatoblasts from pluripotent stem cells. Arch. Toxicol. *91*, 3633–3643.

Meseguer-Ripolles, J., Lucendo-Villarin, B., Wang, Y., and Hay, D.C. (2018). Semi-automated production of hepatocyte like cells from pluripotent stem cells. J. Vis. Exp. *137*, e57995.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Nicola, J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378–1381.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Nekludova, L., Rolfe, P.A., Danford, T.W., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A. (2006). Core transcriptional regulatory circuitry in human hepatocytes. Mol. Syst. Biol. *2*, 1–5.

Parviz, F., Matullo, C., Garrison, W.D., Savatski, L., Adamson, J.W., Ning, G., Kaestner, K.H., Rossi, J.M., Zaret, K.S., and Duncan, S.A. (2003). Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. Nat. Genet. *34*, 292–296.

Patti, G.J., Yanes, O., and Siuzdak, G. (2012). Metabolomics: the apogee of the omics trilogy. Nat. Rev. Mol. Cell Biol. *13*, 263–269.

Ramamoorthy, A., Li, L., Gaedigk, A., Bradford, L.D., Benson, E.A., Flockhart, D.A., and Skaar, T.C. (2012). In silico and in vitro identification of microRNAs that regulate hepatic nuclear factor 4α expression. Drug Metab. Dispos. *40*, 726–733.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389.

Rashidi, H., Luu, N.-T., Alwahsh, S.M., Ginai, M., Alhaque, S., Dong, H., Tomaz, R.A., Vernay, B., Vigneswara, V., Hallett, J.M., et al. (2018). 3D human liver tissue from pluripotent stem cells displays stable phenotype in vitro and supports compromised liver function in vivo. Arch. Toxicol. *92*, 3117–3129.

Si-Tayeb, K., Noto, F.K., Nagaoka, M., Li, J., Battle, M.A., Duris, C., North, P.E., Dalton, S., and Duncan, S.A. (2010). Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. Hepatology *51*, 297–305.

Sladek, F.M., Zhong, W.M., Lai, E., and Darnell, J.E., Jr. (1990). Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. Genes Dev. *4*, 2353–2365.

Stoffel, M., and Duncan, S.A. (1997). The maturity-onset diabetes of the young (MODY1) transcription factor HNF4alpha regulates expression of genes required for glucose transport and metabolism. Proc. Natl. Acad. Sci. U S A *94*, 13209–13214.

Sullivan, G.J., Hay, D.C., Park, I.H., Fletcher, J., Hannoun, Z., Payne, C.M., Dalgetty, D., Black, J.R., Ross, J.A., Samuel, K., et al. (2010). Generation of functional human hepatic endoderm from human induced pluripotent stem cells. Hepatology *51*, 329–335.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Szkolnicka, D., Farnworth, S., Lucendo-Villarin, B., and Hay, D.C. (2014). Accurate predictin of drug-induced liver injury using stem cell-derived populations. Stem Cell Transl. Med. *3*, 141–148.

Tyanova, S., Temu, T., Carlson, A., Sinitcyn, P., Mann, M., and Cox, J. (2015). Visualization of LC-MS/MS proteomics data in MaxQuant. Proteomics *15*, 1453–1456.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740.

Viollet, B., Kahn, A., and Raymondjean, M. (1997). Protein kinase A-dependent phosphorylation modulates DNA-binding activity of hepatocyte nuclear factor 4. Mol. Cell. Biol. *17*, 4208–4219.

Wang, Y., Alhaque, S., Cameron, K., Meseguer-Ripolles, J., Lucendo-Villarin, B., Rashidi, H., and Hay, D.C. (2017). Defined and scalable generation of hepatocyte-like cells from human pluripotent stem cells. J. Vis. Exp. *121*, e55355.

Watt, A.J., Garrison, W.D., and Duncan, S.A. (2003). HNF4: a central regulator of hepatocyte differentiation and function. Hepatology *37*, 1249–1253.

Yokoyama, A., Katsura, S., Ito, R., Hashiba, W., Sekine, H., Fujiki, R., and Kato, S. (2011). Multiple post-translational modifications in hepatocyte nuclear factor 4α. Biochem. Biophys. Res. Commun. *410*, 749–753.

Yusa, K. (2013). Seamless genome editing in human pluripotent stem cells using custom endonuclease-based gene targeting and the piggyBac transposon. Nat. Protoc. *8*, 2061–2078.

Zhou, W., Hannoun, Z., Jaffray, E., Medine, C.N., Black, J.R., Greenhough, S., Zhu, L., Ross, J.A., Forbes, S., Wilmut, I., et al. (2012). SUMOylation of HNF4α regulates protein stability and hepatocyte function. J. Cell Sci. *125*, 3630–3635.

ISCI, Volume *16*

Supplemental Information

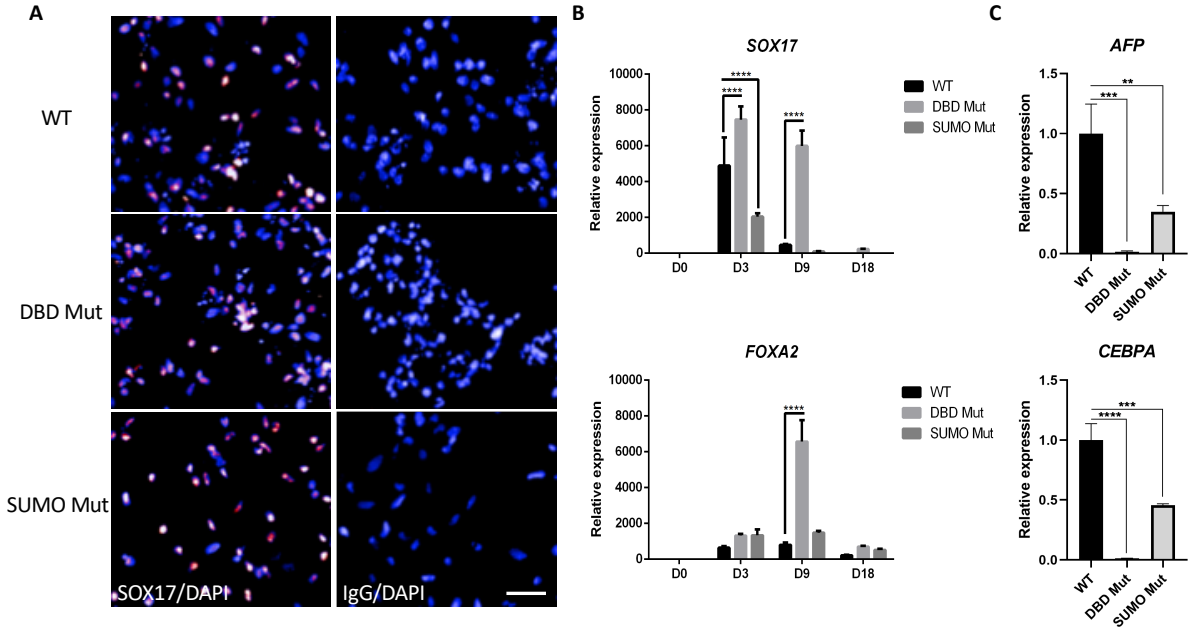
Multiomics Analyses of HNF4α

Protein Domain Function

during Human Pluripotent Stem Cell Differentiation

Yu Wang, Michael H. Tatham, Wolfgang Schmidt-Heck, Carolyn Swann, Karamjit Singh-Dolt, Jose Meseguer-Ripolles, Baltasar Lucendo-Villarin, Tilo Kunath, Timothy R. Rudd, Andrew J.H. Smith, Jan G. Hengstler, Patricio Godoy, Ronald T. Hay, and David C. Hay

SUPPLEMENTAL INFORMATION



Supplemental Figure 1

Figure S1. Related to Figure 2, Differentiation towards the definitive endoderm was not affected in SUMO Mut cells, but was in DBD Mut cells.

(A) Immunostaining of SOX17 in WT, DBD Mut and SUMO Mut pluripotent stem cells differentiated toward definitive endoderm. Scale bar = 50 um.

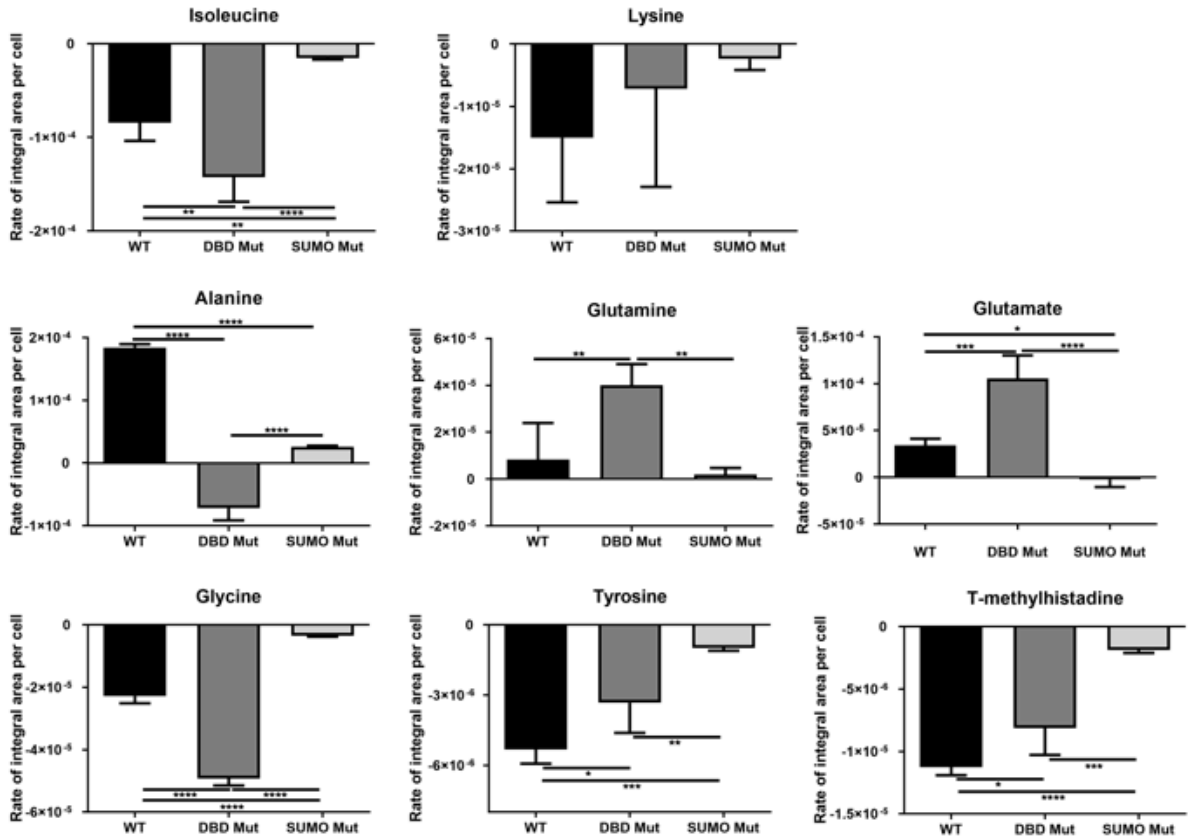
(B) The expression levels of *SOX17* and *FOXA2* during hepatic differentiation. Data were normalized to the housekeeping gene ACTB and expressed relative to day 0 cells. The results shown represent three biological replicates and error bars represent SD. **p < 0.01, *** p < 0.001, **** p < 0.0001; two-way ANOVA with Tukey post hoc test.

(C) Real-time PCR quantification of AFP and CEBPA mRNA level. Data were normalized to the housekeeping gene ACTB and expressed relative to the WT cells. The results shown represent three biological replicates and error bars represent SD. **p < 0.01, *** p < 0.001, **** p < 0.0001; one-way ANOVA with Tukey post hoc test.

Supplemental Figure 2

Figure S2. Related to Figure 2B, Exons 2 and 3 were skipped in HNF4α DBD Mut cells.

Sequencing of HNF4α cDNA in the DBD Mut cells at hepatic progenitor stage showed exon 2 and 3 were skipped.

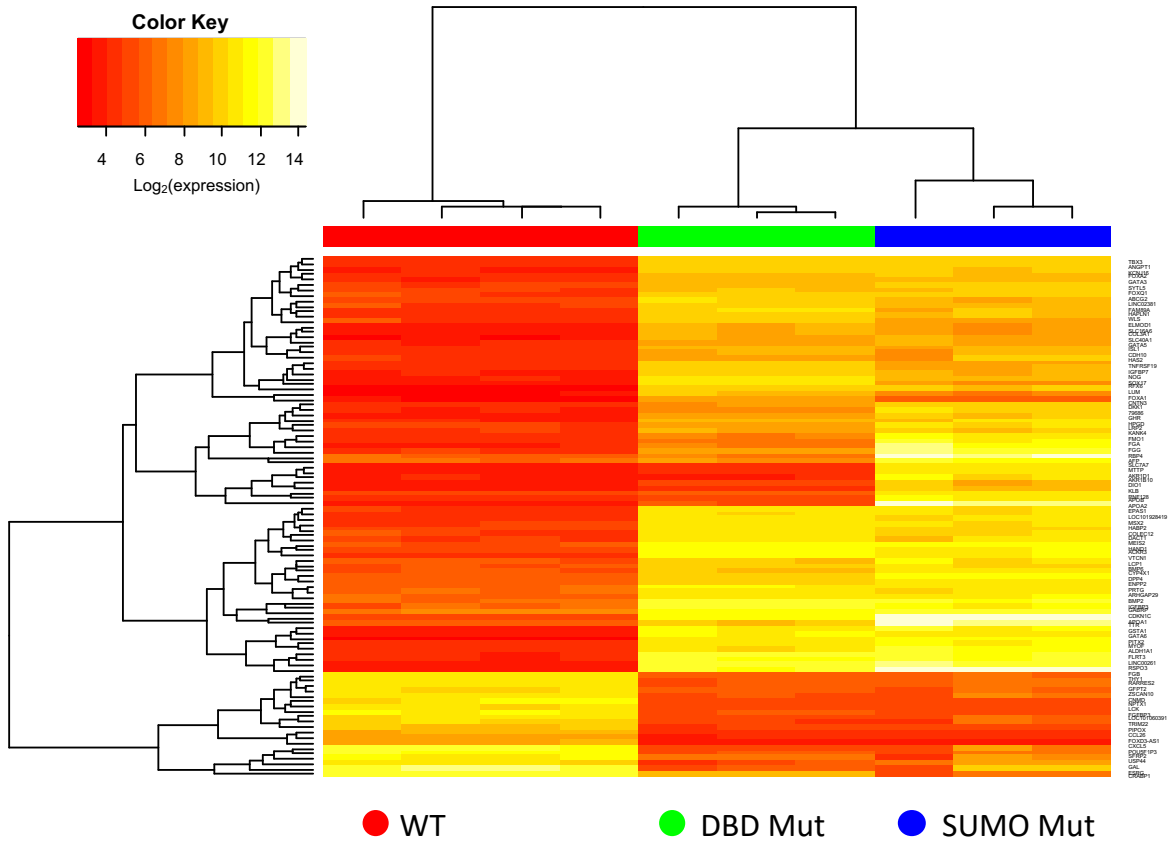


Supplemental Figure 3

Figure S3. Related to Figure 3, Amino acid metabolism was disrupted in the DBD Mut and SUMO Mut hepatoblasts.

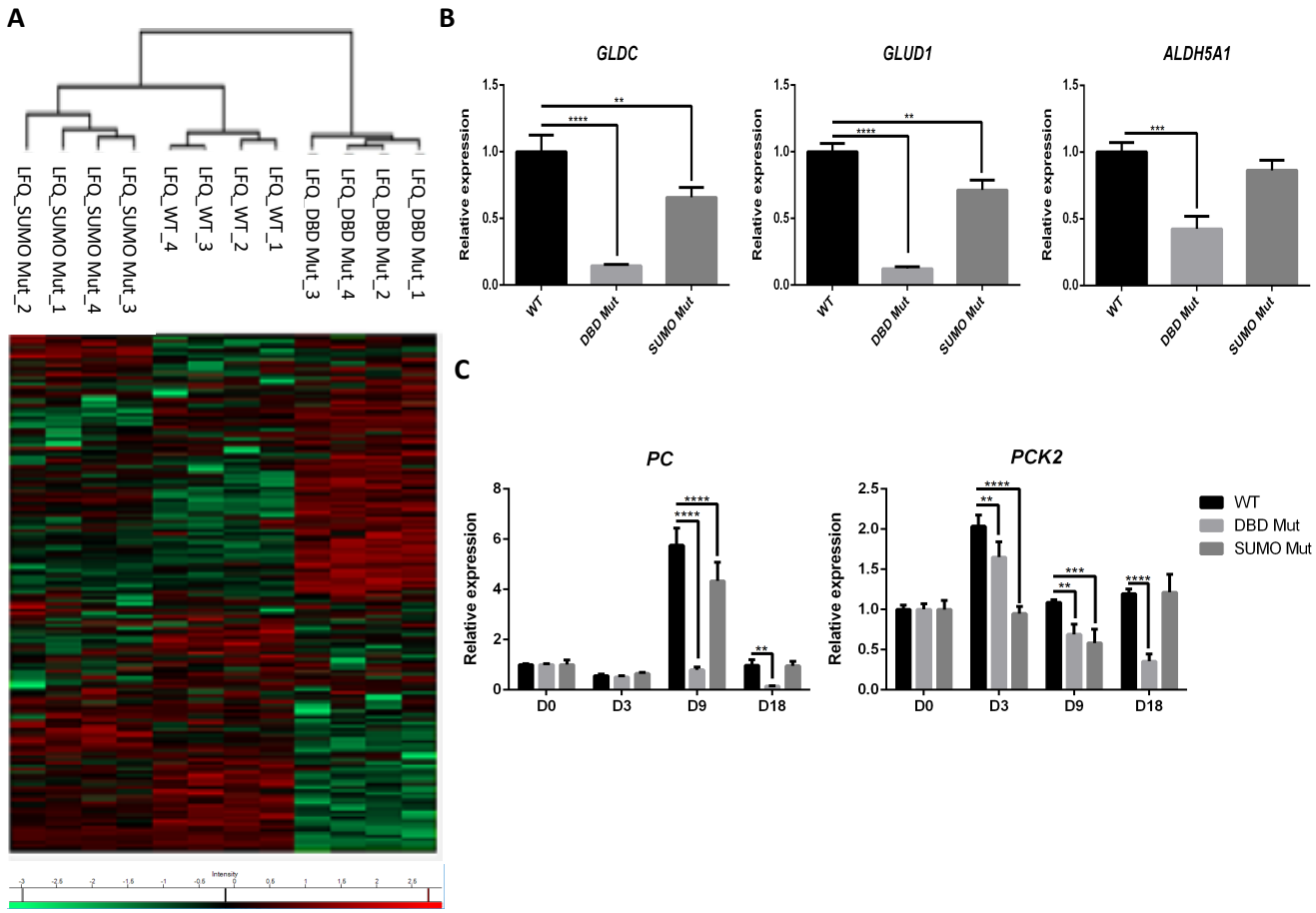
Consumption and production rates of essential amino acids (isoleucine and lysine) and non-essential amino acids (alanine, glutamine, glutamate, glycine, tyrosine and t-methylhistadine) in WT, DBD Mut

and SUMO Mut cells at day 9 in differentiation. Data represent three biological replicates and error bars represent SD. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$; one-way ANOVA with Tukey post hoc test.



Supplemental Figure 4

Figure S4. Related to Figure 4, Heat map representation of the 100 genes with highest variance in WT, DBD Mut and SUMO Mut cells at the hepatic progenitor stage.



Supplemental Figure 5

Figure S5. Related to Figure 5, Heat map representation of the proteomics dataset and the expression of representative metabolic enzymes.

(A) Heat map of proteomics analysis of the WT, DBD Mut and SUMO Mut cells at the hepatic progenitor stage. The LFQ intensities were log10 transformed, then z-scored and normalized by row. (B) The expression of metabolic enzymes GLDC, GLUD1 and ALDH5A1 at mRNA level in the progenitor stage cells. Data were normalized to the housekeeping gene ACTB and expressed relative to the WT cells. The results shown represent three biological replicates and error bars represent SD. **p < 0.01, *** p < 0.001, **** p < 0.0001; one-way ANOVA with Tukey post hoc test. (C) The expression of *PC* and *PCK2* throughout hepatic differentiation process. Data were normalized to the housekeeping gene ACTB and expressed relative to day 0 cells. The results shown represent three biological replicates and error bars represent SD. **p < 0.01, *** p < 0.001, **** p < 0.0001; two-way ANOVA with Tukey post hoc test.

gRNA name	Sequence (5’-3’)
HNF4α-KO-g1-F	caccGCTAGAGAAAGCTGGGGCTCG
HNF4α-KO-g1-R	aaacCGAGCCCCAGCTTTCTCTAGC
HNF4α-KO-g2-F	caccGAACCTGGCCACCTTCTGGGG
HNF4α-KO-g2-R	aaacCCCCAGAAGGTGGCCAGGTTC
HNF4α-KO-g3-F	caccGCCCAGAGACAGCTGCCCTTC
HNF4α-KO-g3-R	aaacGAAGGGCAGCTGTCTCTGGGC
HNF4α-KO-g4-F	caccGTCAGGGAGAAGACAGACCTT
HNF4α-KO-g4-R	aaacAAGGTCTGTCTTCTCCCTGAC
HNF4α-Kin-g11-F	caccGTGCAACAGGTTGTCAATCT
HNF4α-Kin-g11-R	aaacAGATTGACAACCTGTTGCAC
HNF4α-Kin-g12-F	caccGGGAGGTCCGTGCCAAGCCC
HNF4α-Kin-g12-R	aaacGGGCTTGGCACGGACCTCCC
Name	**Sequence (5’-3’)**
5-HA-F	gagctcgtttaaacggcgcgccGCTTGAACCCAGGAGGCA
5-HA-R	gcaatatttcaagaatgcaTAAACTGAATCTGCTCGATCATCTG
3-HA-F	gattatctttctagggttaacGCTCTTCGGCATGGCTAG
3-HA-R	cggcccggccttaattaaTCTCAGGAATCTGCAATTCAC

Table S5. Related to Figure 1, gRNA names and sequences (top panel) and primer names and sequences used in Gibson assembly.

Name	Sequence (5’-3’)
T7E1-KO-F	CCTGGTCCCTTGGCAAGAACT
T7E1-KO-R	AAGACGGGCTCATCTACTGAGGC
T7E1-Kin-F	TTGCCCACCCTCTTCCATTGT
T7E1-Kin-R	CACTTCAGAGAGGTCTCCTGTGGC
Name	**Sequence (5’-3’)**
targeting-LA-F1	CTATGCATGGTGCTAGGTGG
targeting-LA-R1	CATTTTGACTCACGCGGTCG
targeting-LA-R2	TCTCAGGAATCTGCAATTCACTC
targeting-RA-F1	CAGGTGCAGGTGAGCTTGGAGG
targeting-RA-F2	CAGACCGATAAAACACATGCG
targeting-RA-R1	GGATCTCTGAGGTCACCCCTCACTAC

Table S6. Related to Figure 1, Primer names and sequences (top panel) used in T7EI assay and primer names and sequences used in screening of knock-in clones.

Immunostaining Antibody Name	Host	Dilution	Supplier
Primary antibodies			

OCT4	Rabbit Poly	1/200	Abcam
NANOG	Goat Poly	1/200	Abcam
AFP	Mouse Mono	1/500	Abcam
SOX17	Goat Poly	1/500	R & D Systems
HNF4α	Rabbit Poly	1/100	Santa Cruz
IgG	Mouse	1/400	DAKO
IgG	Rabbit	1/400	DAKO
IgG	Goat	1/400	DAKO
Secondary antibodies			
Anti-Mouse 488	Rabbit	1/400	Life Technologies
Anti-Mouse 568	Goat	1/400	Life Technologies
Anti-Rabbit 488	Donkey	1/400	Life Technologies
Anti-Rabbit 568	Donkey	1/400	Life Technologies
Anti-Goat 488	Rabbit	1/400	Life Technologies
Anti-Goat 568	Rabbit	1/400	Life Technologies

Flow Antibody	**Conjugated**	**Type**	**Dilution**	**Supplier**
SSEA-1	FITC	Mouse IgM	1/50	Biolegend
SSEA-4	PE	Mouse IgG	1/50	Biolegend
TRA-1-60	PE	Mouse IgM	1/50	Biolegend

Western Blot Antibody	**Host**	**Dilution**	**Supplier**
Primary antibodies			
HNF4α	Rabbit Ploy	1/1000	Santa Cruz
b-actin	Mouse Mono	1/10000	Sigma
Secondary antibodies			
Anti-Rabbit IgG HRP	Swine	1/3000	R & D Systems
Anti-Mouse IgG HRP	Goat	1/3000	R & D Systems

Table S7. Related to Figure 1 and 2, Optimised antibody concentrations used in immunostaining, flow cytometry and western blotting.

Gene name	Probe number
HNF4α	Hs01023298_m1
HNF1α	Hs00167041_m1
TTR	Hs00174914_m1
APOA2	Hs00155788_m1
APOA4	Hs00166636_m1
HMGCS1	Hs00266810_m1
HMGCS2	Hs00985427_m1
LAMB3	Hs00165078_m1
FGA	Hs00241027_m1
FOXA2	Hs00232764_m1
SOX17	Hs00751752_s1
AFP	Hs01040607_m1
CEBPA	Hs00269972_s1
GLDC	Hs01580591_m1
GLUD1	Hs03989560_s1
ALDH5A1	Hs00542449_m1
PC	Hs00559398_m1
PCK2	Hs00356436_m1
ACTB	Hs01060665_g1

Table S8. Related to Figure 2 and 4, TaqMan probes used in qPCR.

TRANSPARENT METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENTS AND SOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Cell lines
  - Cell Culture Conditions
  - Hepatocyte Differentiation System

- **METHOD DETAILS**
  - CRISPR/Cas9n-sgRNA expression plasmids and targeting vector construction
  - Nucleofection of H9 cells
  - Selection of the nucleofected cells
  - Genotyping the single cell-derived colonies
  - Removal of the transposon in the HNF4α point mutation knock-in experiment
  - Immunostaining
  - Flow cytometry
  - Western blotting
  - ELISA quantification
  - CYP3A assay
  - ^{1}H NMR-based metabolomics analysis
  - Microarray analysis
  - Label-free LC-MS/MS proteomics analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Metabolomics data analysis
  - Microarray data processing and visualization
  - Gene set enrichment analysis (Gene Ontology)
  - Label-free proteomics data analysis
  - Visualization of over-layered analysis of microarray and proteomics datasets
- **DATA AND SOFTWARE AVAILABILITY**

KEY RESOURCES TABLE

REAGENTS or RESOURCES	**SOURCE**	**IDENTIFIER**
Antibodies		
Human NANOG antibody	R&D systems	AF1997
Human OCT4 antibody	Abcam	AB19857
Human SOX17 antibody	R&D systems	AF1924
Anti-Human SSEA4 PE	eBioscience	12-8843-42

Anti-Human CD15(SSEA1) FITC	eBioscience	11-0159-42
Anti-Human TRA-1-60 PE	eBioscience	12-8863-82
HNF4α (H-171)	SantaCruz	Sc-8987
Monoclonal anti-AFP antibody	Sigma	A8452
Monoclonal anti-β-ACTIN	R&D systems	A5441
Mouse IgM isotype control FITC	eBioscience	11-4752-80
Mouse IgG3 isotype control PE	eBioscience	12-4742-41
Polyclonal Swine Anti-Rabbit Immunoglobulins HRP	DAKO	P0399
Polyclonal Goat Anti-Mouse Immunoglobulins HRP	DAKO	P0447
Donkey anti-Goat IgG (H+L) Cross-Adsorbed, Alexa Fluor® 568	Thermo Fisher Scientific	A11057
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed, Alexa Fluor® 488	Thermo Fisher Scientific	A11008
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed, Alexa Fluor® 488	Thermo Fisher Scientific	A21202
4',6-Diamidine-2'-phenylindole dihydrochloride (DAPI)	Roche	10236276001
Mouse IgG	Vector Laboratories	I-2000
Rabbit IgG	Vector Laboratories	I-1000

Goat IgG	Vector Laboratories	I-5000
Bacterial and Virus Strains		
One Shot TOP10 Chemically Competent *E. coli*	Thermo Fisher Scientific	C404003
One Shot Stbl3 chemically competent *E.coli*	Thermo Fisher Scientific	C7373-03
Plasmid Constructs		
pSpCas9n(BB)-2A-Puro (PX462)	Addgene	PX462
pMCS-AAT_PB-PGKpuroTK	Wellcome Trust Sanger Institute	N/A
pCMV-hyPBase	Wellcome Trust Sanger Institute	N/A
pMD19-T vector	Takara	3271
Chemicals, Peptides, and Recombinant Proteins		
Matrigel matrix	Corning	354248
Human Recombinant Laminin 521	BioLamina	LN521
Human Recombinant Laminin 111	BioLamina	LN111
Recombinant Human/Murine/ Rat Activin A (*E.coli* derived)	Peprotech	120-14E
Recombinant Mouse Wnt-3a Protein	R&D systems	1324-WN/CF
Recombinant Human HGF	Peprotech	100-39
Recombinant Human Oncostatin M	Peprotech	300-10

B27 minus vitamin A	Thermo Fisher Scientific	12587-010
RPMI 1640	Thermo Fisher Scientific	11875-093
Penicillin-Streptomycin	Thermo Fisher Scientific	15140-122
Knockout DMEM	Thermo Fisher Scientific	10829-018
Knockout serum replacement	Thermo Fisher Scientific	10828-028
GlutaMAX	Thermo Fisher Scientific	35050-038
NEAA	Thermo Fisher Scientific	11140-035
DMSO	Honeywell Chemicals	D5879
Beta-mecaptoethanol	Thermo Fisher Scientific	31350-010
HepatoZYME-SFM	Thermo Fisher Scientific	17705-021
Hydrocortisone-21 hemisuccinate sodium salt	Sigma	H4881
mTeSR1 medium	Stem Cell Technologies	85850
Gentle Cell Dissociation Reagent	Stem Cell Technologies	07174
TrypLE Express Enzyme	Thermo Fisher Scientific	12604013
Y-27632 (Dihydrochloride)	Millipore	SCM075
DPBS	Thermo Fisher Scientific	14190144

DPBS with calcium and magnesium	Thermo Fisher Scientific	14040133
Bovine Serum Albumin	Sigma	A7906-10G
OneComp eBeads 100 tests	eBioscience	01-1111-42
Trypan Blue Solution, 0.4%	Thermo Fisher Scientific	15250061
NuPAGE 4-12% Bis-Tris Protein Gels, 1.5mm, 10-well	Thermo Fisher Scientific	NP0335BOX
NuPAGE MES SDS Running Buffer (20x)	Life Technologies	NP0002
NuPAGE Transfer Buffer (20x)	Life Technologies	NP0006-1
NuPAGE LDS Sample Buffer	Thermo Fisher Scientific	NP0007
SeeBlue Plus 2 Prestained Standard	Thermo Fisher Scientific	LC5925
TWEEN 20	Sigma	T2700-100ML
Phosphate buffered saline tablet	Sigma	P4417-100TAB
SYBR Safe DNA gel stain	Life Technologies	S33102
UltraPure Agarose	Life Technologies	16500-100
DNA ladder, 1 kb	New England BioLabs	N3232S
UltraPure DNase/RNase-free distilled water	Life Technologies	10977-023
Ampicillin	Sigma	A9518-5G
SOC medium	New England BioLabs	B9020S
LB medium	Sigma	L3022
LB agar medium	Sigma	L2897
FIAU	Moravek	M251

Puromycin dihydrochloride	Life Technologies	A11138-03
FastDigest BbsI (BpiI)	Fermentas/Thermo Scientific	FD1014
DTT	Fermentas/Thermo Scientific	R0862
T4 DNA ligase and buffer	New England BioLabs	M0202S
T4 polynucleotide kinase	New England BioLabs	M0201S
Adenosine 5'-triphosphate, 10 mM	New England BioLabs	P0756S
Q5 High-Fidelity DNA Polymerases	New England BioLabs	M0491
T7 endonuclease I (T7EI)	New England BioLabs	M0302S
NsiI-HF	New England BioLabs	R3127S
AscI	New England BioLabs	R0558S
HpaI	New England BioLabs	R0105S
PacI	New England BioLabs	R0547S
TRIzol Reagent	Thermo Fisher Scientific	15596026
Critical Commercial Assays		
Human stem cell Nucleofector kit 1	Lonza	VPH-5012
P450-Glo™ CYP3A4 Assay System	Promega	V8902
Human Alpha Fetoprotein (AFP) ELISA Kit, 96 tests	Alpha Diagnostic	0500

Human Serum Albumin ELISA Kit, 96 tests	Alpha Diagnostic	1190
GenElute Mammalian Genomic DNA miniprep Kit	Sigma	G1N70-1KT
Pierce BCA Protein Assay Kit	Thermo Fisher Scientific	23225
SuperSignal West Pico Chemiluminescent Substrate	Thermo Fisher Scientific	34077
QIAprep spin miniprep kit	Qiagen	27106
QIAquick gel extraction kit	Qiagen	28704
GeneChip Human Genome HG-U133 Plus 2.0 Array	Affymetrix	N/A
Experimental Models: Cell Lines		
H9 female human ESCs	WiCell	WA09
H9 with HNF4α truncated	This study	N/A
H9 with HNF4α point mutated	This study	N/A
Oligonucleotides		
See Table S5, S6	N/A	N/A
Recombinant DNA		
gRNA for CRISPR cloned in SpCas9n(px462), see Table S5 for oligonucleotide sequences	This study	N/A
PB_HNF4α_PGK puroTK	This study	N/A
Software and Algorithms		

Flowjo	FLOWJO, LLC	https://www.flowjo.com/
FACSDIVA software	BD Biosciences	N/A
GraphPad Prism	GraphPad Software	https://www.graphpad.com/scientific software/prism/
Columbus Image Data Storage and Analysis System	PerkinElmer	http://www.cambridgesoft.com/ensemble/spotfire/Columbus/default.aspx
Chenomx NMR Suite version 8.1	Chenomx	https://www.chenomx.com/software/archive/
Microarray Suite software	Affymetrix	N/A
Limma packages	Bioconductor	https://bioconductor.org/packages/release/bioc/html/limma.html
MaxQuant proteomics software	Max Planck Institute of Biochemistry	http://www.biochem.mpg.de/5111795/maxquant
Perseus proteomics software	Max Planck Institute of Biochemistry	http://www.biochem.mpg.de/5111810/perseus
Enrichr analysis website	Icahn School of Medicine at Mount Sinai	http://amp.pharm.mssm.edu/Enrichr/
STRING protein interaction analysis website	N/A	https://string-db.org
PaintOmics3	N/A	http://bioinfo.cipf.es/paintomics/

CONTACT FOR REAGENTS AND SOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David C. Hay (davehay@talktalk.net).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

H9 female human ESCs were purchased from WiCell. Cells were routinely checked for cell surface markers using flow cytometry and for the presence of mycoplasma.

Cell Culture Conditions

H9 human ESCs and genome modified DBD Mut and SUMO Mut cell lines were cultured on Matrigel coated cell culture plates with mTeSR1 medium. Cells were routinely split in 1:3 ratio

using Gentle Cell Dissociation reagent. For nucleofection, single cells were obtained by treating the cells with Gentle Cell Dissociation reagent for 7-8 minutes at 37 ^{o}C. ROCK inhibitor was used to maintain the single cells.

Hepatocyte Differentiation System

Hepatocyte differentiation of the different cell lines was carried out following lab-developed protocol (Wang et al., 2017). Briefly, 5.5-6.5x10^5 single cells were seeded into per well on 6-well plates coated with LN521/LN111 mix (1:3) for differentiation. Once cells reached approximately 40% confluency post seeding, hepatocyte differentiation was initiated by switching the mTeSR1 medium to RPMI 1640 medium supplemented with 2 % B27 (minus Vitamin A), 100 ng/ml Activin A and 50 ng/ml Wnt 3a. This definitive endoderm induction medium was changed daily for 3 days. Subsequently, the differentiation medium was switched to Knockout DMEM medium supplemented with 20 % knockout serum replacement, 0.5 % GlutaMAX, 1% NEAA and DMSO, and 0.1 mM beta mecaptoethanol. This hepatic progenitor specification medium was changed daily for 5 days. Following this, the differentiation medium was changed to HepatoZYME medium supplemented with 1 % GlutaMAX, 10 uM hydrocortisone, 20 ng/ml hepatocyte growth factor (HGF) and 10 ng/ml oncostatin M (OSM). This hepatocyte-like cells maturation medium was changed every other day for 10 days. The derived hepatocyte-like cells were checked for functions, including basal cytochrome P450 3A activity, albumin and AFP secretion. In addition, cells at representative time points (day 0, day 3, day 9 and day 18) were routinely checked for the expression of typical markers corresponding to that time point.

METHOD DETAILS

CRISPR/Cas9n-sgRNA expression plasmids and targeting vector construction

Oligonucleotides for gRNAs (Table S5) targeting HNF4α were annealed and cloned into a BbsI digested Cas9n backbone plasmid (PX462) following a published protocol (Ran et al., 2013).

The construction of the *piggyBac*-based targeting construct involved three main steps: the amplification of the 5'-homology arm (HA) and the 3'-HA; the insertion of the 5'-HA into the backbone plasmid; and the insertion of the 3'-HA into the backbone plasmid.

The PCR primers for amplifying 5'-HA and 3'-HA used in Gibson assembly were generated using NEBuilder (https://nebuilder.neb.com) (Table S5). The desired point mutations were incorporated into the primers. The PCR amplification of the two homology arms were catalysed by Q5 high-fidelity DNA polymerase (New England Biolabs). H9 genomic DNA was used as the template for the PCR. The amplified homology arms were then purified before being inserted into the backbone plasmid.

The used *piggyBac* backbone was a kind gift of Dr. Kosuke Yusa from the Wellcome Trust Sanger Institute, UK. For the insertion of the 5'-HA, the backbone plasmid was digested using NsiI and AscI enzymes and then gel purified. The purified 5'-HA was then inserted using a Gibson assembly reaction. The *piggyBac*-5'-HA plasmid was then purified and sequenced to confirm the correct insertion of the 5'-HA. Following this, the *piggyBac*-5'-HA plasmid was digested using HpaI and PacI enzymes and then purified, the 3'-HA was then inserted into the

backbone and completed the construction of the donor plasmid. Sequencing was also performed to confirm the correct insertion of the 3'-HA.

Nucleofection of H9 cells

H9 cells at 70-80 % confluency were dissociated using Gentle Cell Dissociation reagent to make single cell suspension. For HNF4α gene knockout, ~ 0.8 million live cells were nucleofected with 3 μg of each Cas9n-sgRNA expression plasmid. For the knockin experiment, ~ 0.8 million cells were nucleofected with 5 μg donor plasmid and 3 μg of both Cas9n-sgRNA expression plasmids. The Human Stem Cell Nucleofector® Kit 1 (Lonza) was used and the nucleofection was facilitated by a Nucleofector 2b device with program A-023 (Lonza).

Selection of the nucleofected cells

48 hours post nucleofection, the cells were selected using 0.5 μg/ml puromycin for 2 days. Puromycin-resistant cells were then propagated and passaged to 10-15 96-well plates at the concentration of 0.8 cell/well in mTeSR1 medium supplemented with 10 μM Rho-associated protein kinase (ROCK) inhibitor (Sigma-Aldrich). At this point, puromycin in the medium for HNF4α knockout experiment cells was removed, but it was kept for selecting cells for HNF4α point mutation knock-in experiment. Following the seeding, the cells were maintained at 37 ^{o}C with 10 % CO2 for 7 days to form single cell-derived colonies.

The single-cell derived colonies were picked at about another 5 days later and passaged at a 1:2 ratio to two different plates. ~4 days after the passaging, cells on one plate were lysed for genomic DNA extraction.

Genotyping the single cell-derived colonies

Colonies from HNF4α knockout experiments were analysed using T7EI assay (Table S6). The targeted region was PCR amplified and the PCR products were then digested using T7EI enzyme, and subsequently separated on agarose gels and the genotypes were 'readable' on the gel.

Colonies from HNF4α point mutation knock-in experiments were amplified using three-primer PCRs (Table S6) to differentiate genomic sequences and inserted sequences. The PCR products were separated on agarose gels and the genotype were 'readable' on the gel.

For promising colonies from the DNA gel-based genotyping, the PCR products were purified and then inserted into a pMD19-T vector to form a stable plasmid for sequencing. The colonies with correct genotypes were then saved and expanded to form stable isogenic cell lines.

Removal of the transposon in the HNF4α point mutation knock-in experiment

The correctly targeted colonies were established and two were used for the removal of the selection cassette introduced by the transposon. Briefly, ~0.8 million cells were nucleofected with 5 μg hyperactive PBase (gift from Dr. Kosuke Yusa) and the cells were culture in mTeSR1 medium without puromycin. The cells were then maintained for 3-4 passages before the negative selection mediated by the thymidine analog 1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-5-iodouracil (FIAU) at a concentration of 200 nM. About 2 weeks after the

negative selection, single cell-derived colonies were picked and genotyped as described above.

Immunostaining

Cells were fixed in 100 % ice-cold methanol at -20 ^{o}C for 30 minutes. Post fixation, cells were washed three times with DPBS for 5 minutes at room temperature. The cells were subsequently blocked with PBS-0.1 % Tween (PBST) containing 10 % BSA for 1 hour at room temperature. After that, the blocking solution was removed and the respective primary antibody diluted in 1% BSA/PBST was added to the cells. The primary antibody was incubated at 4 ^{o}C overnight with gentle agitation. Following this, the used primary antibody solution was removed and the cells were washed three times with PBST for 5 minutes at room temperature. The cells were then incubated with the appropriate secondary antibody diluted in PBST for 1 hours at room temperature in the dark with gentle agitation. The cells were then washed three times with PBST. Following this, cells were stained with DAPI diluted in PBS (1:5000) for 2 minutes at room temperature. The DAPI solution was then removed and the cells were washed twice with PBST. In all cases, the stained cells were stored at 4 ^{o}C in the dark before imaging. The primary and secondary antibodies are listed in supplemental Table S7.

All images were collected at room temperature using the automated Operetta fluorescent microscope. The images were processed using Columbus Image analysis server.

Flow cytometry

Fluorescence activated cell sorting (FACS) was used to confirm the expression of cell surface markers in hESCs. hESCs from a well on a 6-well plate were washed once with 2 ml DPBS and then dissociated with 1ml TrypLE. The cells were lifted off as single cells and collected. Post centrifugation, cells were resuspended in DPBS and filtered through a 0.22 μm filter. The single cell suspension was then incubated with the fluorochrome conjugated antibodies and the corresponding IgG controls for 30 minutes at 4 ^{o}C. The antibody information is listed in supplemental Table S7. Cells were then washed twice with DPBS, removing any unbound antibody, and span down at 200 x g for 5 minutes. Following this, cells were resuspended in 300 μl of DPBS.

Dead cells and debris were not included in the analysis. This was carried out by using an electronic live gate on forward scatter and side scatter parameters. Data for 20,000-50,000 'live' events were acquired for each sample using a BD LSR Fortessa (4 laser) analyser. The data was analysed using FlowJo software.

Western blotting

Cells at different time points were washed once with DPBS and lysed in wells at room temperature in dark for 5 min. 200 μl home-made SUMO lysis buffer (2 % SDS, 50 mM Tris-HCL (pH 8.0), 1 mM EDTA, 10 mM iodoacetamide) supplemented with proteinase and phosphatase inhibitors (Sigma-Aldrich) at 1 % final concentration was used for each well of a 6-well plate. The cell extract was briefly sonicated using a Bioruptor (Diagenode) and span down for 15 min at 12,000 rpm at 4 ^{o}C. The protein supernatant was then transferred to a new 1.5 ml Eppendorf tube.

The Pierce bicinchoninic acid (BCA) assay kit (Thermo Fisher Scientific) was used to quantify the protein concentration. Protein supernatant was diluted 1:10 using the lysis buffer. 20 μl protein supernatant of each sample was analysed in duplicates. Reagent A and B from the kit were mixed at a 1:50 ratio and a volume of 200 μl was transferred into each sample well. A standard curve was generated using the bovine serum albumin standards at concentrations ranging from 0-2000 μg/ml. The plate was incubated at room temperature for 30 min and the absorbance was read at 562 nm. The protein concentrations were calculated by linear extrapolation using the standard curve generated.

The SDS-NuPAGE polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins of different sizes. 50 μg protein supernatant was denatured at 100 ^{o}C for 10min in 1x NuPAGE LDS sample buffer supplemented with 10 mM DTT before being used for electrophoresis. 4-12 % Bis-Tris precast polyacrylamide gels (Thermo Fisher Scientific) were used with the XCell SureLock Mini-Cell System (Thermo Fisher Scientific) for the electrophoresis. Once the gel was fitted in the chamber, the tank was filled with 1x NuPAGE MES-SDS running buffer with 0.5 ml of NuPAGE Antioxidant (Thermo Fisher Scientific) added into the inner chamber. The samples were loaded along with a SeeBlue Plus 2 Pre-Stained Standard (Invitrogen). A current of 200 V was applied and the samples were run for 1-2 hours depending on the purpose of each experiment. The gels containing proteins were carefully removed from the cassette and were used in western blotting.

Following the protein separation using SDS-PAGE gels, the proteins were then transferred from the gels to the Polyvinylidene fluoride (PVDF) membrane. The transfer sack was assembled in the following order from cathode to anode: 3x sponge; 2x filter paper soaked in 1x NuPAGE Transfer Buffer; SDS-PAGE gel; PVDF membrane activated in 100 % ice-cold methanol; 2x filter paper soaked in 1x NuPAGE Transfer Buffer; 3x sponge. No bubbles between the gel and the membrane were allowed. Once the sack was assembled and placed into the XCell Blot II module, the module was then tightly sealed and placed into the transfer SureLock tank containing 1x transfer buffer in the inner chamber and ice-cold water in the outer chamber. The tank was then placed on ice for the entire protein transfer process. A constant current of 160mA was applied for 90-120 min.

Once the proteins were successfully transfer onto the PVDF membrane, the membrane was blocked in 10 % skimmed milk at room temperature with gentle agitation for at least one hour to prevent non-specific antibody binding.

Following this, the membrane was probed with desired primary antibody diluted in 5 ml 10 % BSA/PBST at 4 ^{o}C overnight with gentle rolling. The dilution ratios were optimised specifically for each primary antibody and were listed in Table 4. Unbound antibody was removed by three, 5-min washes with adequate 0.1 % PBST at room temperature with agitation. The primary antibody could be re-used for 2-3 times if stored at 4 ^{o}C properly.

After probed with the primary antibody, the membrane was then incubated with corresponding horseradish peroxidase (HRP)-conjugated secondary antibody diluted in 5 ml 10 % skimmed milk at room temperature for at least one hour with gentle rolling. Unbound secondary antibody was then removed by three, 5-min washes with adequate 0.1 % PBST at

room temperature with agitation. The primary and secondary antibody information is listed in supplemental Table S7.

The proteins of interest were then detected using enhanced chemiluminescence (ECL). Proteins bands were visualised using the Pierce Enhanced Chemiluminescence Kit (Pierce, UK). Peroxidase buffer and the Luminol/Enhancer solution was mixed at a 1:1 ratio and spread evenly onto the membrane (2 ml for each membrane), followed by a 5-min incubation at room temperature. The HRP substrate reacts with the conjugated HRP group presented on the secondary antibody, specifying the target protein. The membrane was developed in the dark room using a film developer.

ELISA quantification

At the hepatocyte stage, day 18 cell culture supernatants were collected. Albumin and AFP concentration were determined using commercially available kits (Alpha Diagnostic Intl. Inc, San Antonio, USA). The ELISA assay was performed according to the manufacturer's instructions with cell culture supernatants diluted at 1:3 (albumin) or 1:10 (AFP). After one hour incubation at room temperature, the wells were washed for four times and anti-human albumin or AFP HRP-conjugated secondary antibodies were added to the wells for 30 minutes. Subsequently, the wells were washed five times and the substrate was then added and incubated for a further 15 minutes in the dark at room temperature. Following this, the stop solution was added directly to the wells in order to stop the enzymatic reaction. The plates were then read at 450 nm with a reference wavelength of 630nm using a FLUOstart Omega plate reader (BMG LabTech, Germany). Blank cell culture media incubated at 37 ^{o}C for 48 hours, diluted in 1:3 or 1:10 was used as a negative control. The data was then normalised to per ml per 48 hours per mg protein as determined by BCA assay (Pierce, UK).

Quantitative real-time PCR

Total RNA was collected from cells using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Total RNA was reverse transcribed using QuantiTect reverse-transcription kit (Qiagen) following the manufacturer's protocol. Real-time PCR was performed using TaqMan Fast Universal Master Mix and appropriate Taqman probes (Applied Biosystems). The primers are listed in supplemental Table S8. The samples were analysed using Roche LightCycler 480 Real-Time PCR System. Results were normalized to beta-actin. The qPCRs were run in triplicates.

CYP3A assay

Day 18 hESC-derived hepatocyte-like cells (HLCs) were incubated with the luciferin conjugated CYP3A (1:40) substrate (P450 P-Glo Luminescent Kit, Promega, UK) diluted in the maturation medium for 5 hours at 37 ^{o}C. Blank cell culture medium was used as a negative control. The supernatants were then collected and tested for luminescence. The luciferin detection reagent was reconstituted by mixing the buffer with the lyophilised luciferin detection reagent. 50 μl of the collected supernatant was mixed with 50 μl of the detection reagent in a white 96 well plate and incubated at room temperature in the dark for 20 minutes. The relative levels of basal activity were measured using a luminometer (POLARstar optima). Units of activity were recorded as relative light units per ml per mg protein (RLU/ml/mg) as normalised by protein content.

^{1}H NMR-based metabolomics analysis

24-hour cell culture medium from cells at differentiation day 9 was collected. The medium was then centrifuged at 4 ^{o}C, 300 x g for 5 min. The supernatant was collected and flash-frozen on dry ice.

All media samples were analysed on a 700 MHz Bruker Avance Neo NMR spectrometer equipped with a BBI room temperature probe. The aliquots of the conditioned media from each cell line and unconditioned controls were thawed at room temperature and 550 μL of each was transferred into a microcentrifuge tube, to this 50 μL of deuterium oxide containing 1 mM deuterated 3-(Trimethylsilyl)-1-propanesulfonic acid-d6 sodium (DSS-d6, 98% atom %D, Sigma-Aldrich), which was used as an internal reference for calibration, was added. The total volume of 600 μL was transferred into individually coded sample tubes in 96-racks (Bruker).

Briefly, the samples were loaded using a SampleJet sample changer. The samples remained chilled at 277 K on the sample changer and at 300 K during acquisition of the ^{1}H NMR experiments. Matching, tuning and shimming were performed automatically on each individual sample to reduce sample-to-sample variation. ^{1}H NMR spectra were collected using a standard 1D-nuclear overhauser enhancement spectroscopy (NOESY)-presaturation and THE PROJECT-Carr-Purcell-Meiboom-Gill (CPMG) pulse sequences. The RF field strength for presaturation was 25Hz. For both experiments 8 scans were collected with a spectral width of 20 ppm, an acquisition time of 3.0 s and a relaxation delay of 40.0 s, which was determined from T1 measurements. The value of O1 was determined automatically for the first 1D-NOESY experiment and was used subsequently for the CPMG experiment. The total CPMG period was 64 ms with 180^{o} every 400 μs. For both experiments a signal-to-noise ratio of greater than 100:1 was achieved. A range of metabolites were identified with the aid of the software Chenomx NMR Suite version 8.1 (Chenomx Inc, Edmonton, Alberta) and human metabolome database (HMDB).

Microarray analysis

Cells at differentiation day 9 on Laminin 521/111 mix (1:3 ratio) were lysed in TRIZOL and stored at -80 ^{o}C. RNA extraction was performed according to the manufacturer's instructions. Analysis of gene expression was performed using Affymetrix GenChip Human Genome HG-U133 plus 2.0 array (Santa Clara, CA, USA). Microarrays were scanned with an Affymetrix scanner controlled by Affymetrix Microarray Suite software.

Label-free LC-MS/MS proteomics analysis

To facilitate whole proteome analysis of the three lines, quadruplicate samples of WT, DBD Mut and SUMO Mut cells were taken at differentiation day 9. Total protein extract was prepared in 1x LDS sample buffer, and 20 μg total protein for each sample was sonicated and boiled for 10 minutes prior to fractionation by NuPAGE 10% Bis-Tris gels using 3-(N-morpholino) propanesulfonic acid (MOPS) buffer. Each lane of the gel was excised into 4 slices, and peptides were prepared in parallel by in-gel tryptic digestion (Shevchenko et al., 2007). For each sample, approximately 1μg total peptide was submitted for liquid chromatography coupled tandem MS (LC-MS/MS) analysis on a Q Exactive mass spectrometer (Thermo Scientific) coupled to an EASY-nLC 1000 liquid chromatography system (Thermo Scientific). Peptides were fractionated on a 75 μm x 500 mm EASY-Spray column (Thermo

Scientific) using a 90-minute gradient. Precursor ion full scan spectra were acquired over (m/z 300 to 1,800) with a resolution of 70,000 at m/z 200 (target value of 1,000,000 ions, maximum injection time 20 ms). Up to ten data dependent MS2 spectra were acquired with a resolution of 35,000 at m/z 200 (target value of 1,000,000 ions, maximum injection time 120 ms). Ions with unassigned, +1 and +8 charge state were rejected. Intensity threshold was set to 2.1 x 10^4 units. Peptide match was set to preferred, dynamic exclusion duration 40 s.

QUANTIFICATION AND STATISTICAL ANALYSIS

Metabolomics data analysis

The metabolic rate for each metabolite was quantified by normalizing the signal intensity against the cell number at the detected time point. The unit for the metabolic rate is then recorded as rate of integral area per cell. A negative rate indicates consumption, while a positive value means production of the metabolite. All samples were tested in triplicates.

Microarray data processing and visualization

Affymetrix gene expression data were pre-processed using 'affyPLM' packages of the Bioconductor Software (Gentleman et al., 2004). To obtain the genes with the strongest evidence of differential expression, a linear model fit was applied for each gene using 'limma' (Linear Models for Microarray Data) packages of the Bioconductor Software.

Gene expression levels in each HNF4α-edited cell line were log2 transformed and then compared to the wild type control cells. Genes with a fold change greater than two over the wild type cell expression levels (p value <= 0.05, false discovery rate (1% FDR) corrected) were taken as significantly deregulated (Table S1). All samples were tested in triplicates. Processing and visualization (Principal Component Analysis) of data were performed using MATLAB tools (The MathWorks Inc., Natick, MA, USA).

Gene set enrichment analysis (Gene Ontology)

Genes which showed change of equal to or more than 2-fold were subjected to gene ontology (GO) analyses using the web server Enrichr (Chen et al., 2013; Kuleshov et al., 2016). The results of biological pathways (GOBP) were provided in Table S2.

Label-free proteomics data analysis

Data were analyzed using MaxQuant (version 1.5.8.3) (Cox and Mann, 2008; Cox et al., 2011) and searched against UniProtKB *H. sapiens* database (86749 sequences - 13/06/2012). Unless otherwise stated, all MaxQuant settings were default. Variable modifications of acetyl (Protein N-term) and oxidation (M) were considered, along with the fixed modification of carbamidomethyl (C). Enzyme specificity was set to trypsin/P, and a false discovery rate of 1 % was set as a threshold at protein, peptide and site levels, and a mass deviation of 6 ppm was set for main search and 20 ppm for MS2 peaks. Match between runs was applied, and label-free quantification (LFQ) was selected.

The MaxQuant proteinGroups.txt file was used for downstream proteome analysis. All decoy and 'putative contaminant' entries were removed, as were any identified only by site and those with fewer than 4 reported LFQ values for any single experimental group (WT, DBD Mut

and SUMO Mut). This left 3639 protein groups comparable across all cell types. Zero LFQ values were replaced in Perseus (Tyanova et al., 2016), from a normal distribution of existing $\log_{10}$ LFQ intensities downshifted 1.8 and of width 0.3. Proteins were defined as statistically differing between groups using the Perseus unpaired two-samples Student's t-test truncated by 1% permutation-based FDR using an S0 value of 0.1. No apparent specificity was entered as an empty cell in the spreadsheet.

PCA plotting was performed using Perseus software. GOBP enrichment analysis was performed on Enrichr server (Table S4).

Protein-protein interaction visualization was performed using a web server named STRING (https://string-db.org).

Visualization of over-layered analysis of microarray and proteomics datasets

A free web-based multi-omics data visualization application named PaintOmics 3 (http://bioinfo.cipf.es/paintomics/) (Hernandez-de-Diego et al., 2018) was used to do the over-layered analysis of microarray and proteomics datasets.

DATA AND SOFTWARE AVAILABILITY

Data availability

The accession number for the microarray data reported in this paper is EMBL-EBI: E-MTAB-7951.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD013737.

REFERENCE

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J. V., and Mann, M. (2011). Andromeda: A peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794–1805.

Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Hernandez-de-Diego, R., Tarazona, S., Martinez-Mira, C., Balzano-Nogueira, L., Furio-Tari, P., Pappas, G.J., and Conesa, A. (2018). PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. BioRxiv 1–7.

Kuleshov, M. V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.
Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: Improving support for quantification data. Nucleic Acids Res. *47*, D442–D450.
Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.
Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. V., and Mann, M. (2007). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. *1*, 2856–2860.
Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740.
Wang, Y., Alhaque, S., Cameron, K., Meseguer-Ripolles, J., Lucendo-Villarin, B., Rashidi, H., and Hay, D.C. (2017). Defined and Scalable Generation of Hepatocyte-like Cells from Human Pluripotent Stem Cells. J. Vis. Exp. 1–8.